

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2001, 14:35:56 ; Search time 3871.54 Seconds  
(without alignments)  
11781.976 Million cell updates/sec

Title: US-09-541-094-13  
Perfect score: 2949  
Sequence: 1 tctgcagaattcgcgttcg.....atlaaataagattatccac 2949

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_bal:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: gb\_bal:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vil:\*  
59: gb\_vil2:\*  
60: gb\_vil3:\*  
61: gb\_vil4:\*  
62: gb\_vil5:\*  
63: gb\_vil6:\*  
64: gb\_vil7:\*  
65: gb\_vil8:\*  
66: gb\_vil9:\*  
67: gb\_vil10:\*  
68: gb\_vil11:\*  
69: gb\_vil12:\*  
70: gb\_vil13:\*  
71: gb\_vil14:\*  
72: gb\_vil15:\*  
73: gb\_vil16:\*  
74: gb\_vil17:\*  
75: gb\_vil18:\*  
76: gb\_vil19:\*  
77: gb\_vil20:\*  
78: gb\_vil21:\*  
79: gb\_vil22:\*  
80: gb\_vil23:\*  
81: gb\_vil24:\*  
82: gb\_vil25:\*  
83: gb\_vil26:\*  
84: gb\_vil27:\*  
85: gb\_vil28:\*  
86: gb\_vil29:\*  
87: gb\_vil30:\*  
88: gb\_vil31:\*  
89: gb\_vil32:\*  
90: gb\_vil33:\*  
91: gb\_vil34:\*  
92: gb\_vil35:\*  
93: gb\_vil36:\*  
94: gb\_vil37:\*  
95: gb\_vil38:\*  
96: gb\_vil39:\*  
97: gb\_vil40:\*  
98: em\_ba3:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2948.2	100.0	2949	89	AF240468	AF240468 Homo sapi
2	2804.2	95.1	2805	91	D87442	D87442 Human mRNA
3	1746.4	59.2	2250	94	AF240469	AF240469 Mus muscu
4	802	27.2	155913	81	AL513282	AL513282 Homo sapi
5	802	27.2	182344	81	AL445230	AL445230 Homo sapi
6	791	26.8	170677	78	AL138932	AL138932 Homo sapi
7	332	11.3	165942	91	AP001605	AP001605 Homo sapi
8	332	11.3	340000	91	AP001699	AP001699 Homo sapi

C	9	325.6	11.0	150597	77	AC090058	AC090058 Homo sapi
C	10	262.6	8.9	223431	75	AC074310	AC074310 Mus muscu
C	11	256.2	8.7	190442	66	AC021119	AC021119 Homo sapi
C	12	213.6	7.2	170677	78	AL138932	AL138932 Homo sapi
C	13	180.8	6.1	148540	81	AL391417	AL391417 Homo sapi
C	14	180.8	6.1	167253	90	AL445926	AL445926 Human DNA
C	15	180.8	6.1	168537	82	AL590397	AL590397 Homo sapi
C	16	180.8	6.1	173152	70	AC026466	AC026466 Homo sapi
C	17	180.8	6.1	246646	81	AL391275	AL391275 Homo sapi
C	18	180.8	6.1	302665	81	AL391377	AL391377 Homo sapi
C	19	144.6	4.9	4692	97	HS024105	U24105 Homo sapien
C	20	79.2	2.7	7218	10	166494	166494 Sequence 14
C	21	59.4	2.0	858	53	CNS06CYX	AL416479 T7 end of
C	22	54.4	1.8	2909	5	AF240470	AF240470 Drosophil
C	23	54	1.8	187540	75	AC073970	AC073970 Homo sapi
C	24	53.2	1.8	154608	74	AC073620	AC073620 Homo sapi
C	25	51.4	1.7	72500	77	AC090206	AC090206 Homo sapi
C	26	50.6	1.7	258174	76	AC079429	AC079429 Mus muscu
C	27	50	1.7	101509	70	AC027353	AC027353 Homo sapi
C	28	50	1.7	148600	70	AC027275	AC027275 Homo sapi
C	29	49.8	1.7	53522	97	HUMPKD1GEN	L39891 Homo sapien
C	30	49.8	1.7	53526	9	AR097326	AR097326 Sequence
C	31	49.8	1.7	53577	9	AR097325	AR097325 Sequence
C	32	49.4	1.7	190492	75	AC073957	AC073957 Homo sapi
C	33	49.4	1.7	298166	77	AC087563	AC087563 Homo sapi
C	34	49	1.7	1063	53	CNS06L6W	AL403806 T3 end of
C	35	48.4	1.6	58998	63	AC014104	AC014104 Drosophil
C	36	48.4	1.6	215711	4	AC008205	AC008205 Drosophil
C	37	48.4	1.6	228479	5	AE003749	AE003749 Drosophil
C	38	48.2	1.6	139524	78	AC090444	AC090444 Rattus no
C	39	47.8	1.6	5809	85	AB007974	AB007974 Homo sapi
C	40	47.8	1.6	92775	90	AL441964	AL441964 Human DNA
C	41	47.8	1.6	213432	73	AC068198	AC068198 Homo sapi
C	42	47.8	1.6	267759	76	AC079418	AC079418 Homo sapi
C	43	47.6	1.6	14955	5	AF000580	AF000580 Dictyoste
C	44	47.6	1.6	152513	83	AP003207	AP003207 Oryza sat
C	45	47.6	1.6	165608	77	AC087842	AC087842 Rattus no

## ALIGNMENTS

RESULT 1  
LOCUS AF240468 2949 bp mRNA PRI 08-SEP-2000  
DEFINITION Homo sapiens nicastrin mRNA, complete cds.  
ACCESSION AF240468  
VERSION AF240468.1 GI:9992877  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2949)  
AUTHORS Yu, G., Nishimura, M., Arakawa, S., Levitan, D., Zhang, L., Tandon, A., Song, Y.-Q., Rogaeve, E., Chen, F., Kawarai, T., Supala, A., Levesque, L., Xu, H., Yang, D. S., Holmes, E., Milman, P., Liang, Y., Zhang, D. M., Xu, D. H., Sato, C., Rogaeve, E., Smith, M., Janus, C., Zhang, Y., Aebersold, R., Farrer, L. S., Sorbi, S., Bruni, A., Fraser, P. and St George-Hyslop, P.

TITLE Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing  
JOURNAL Nature 407 (6800), 48-54 (2000)  
MEDLINE 20445163  
REFERENCE 2 (bases 1 to 2949)

AUTHORS Yu, G., Nishimura, M., Arakawa, S., Levitan, D., Zhang, L., Tandon, A., Song, Y.-Q., Rogaeve, E., Kawarai, T., Supala, A., Aebersold, R., Sorbi, S., Bruni, A., Fraser, P. E. and St George-Hyslop, P. H.

TITLE Submitted (02-MAR-2000) Dept. of Medicine, Centre for Research in Neurodegenerative Diseases, University of Toronto and University Health Network, Rm. 118, Tanz Neuroscience Bldg., 6 Queen's Park Crescent West, Toronto, Ontario M5S 3H2, Canada

FEATURES	Location/Qualifiers
source	1..2949
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
	/map="near D12S1595 and D1S2844"
	143..2272
CDS	/function="modulates beta-APP processing and Notch/glp1 signal transduction"
	/note="type 1 transmembrane protein; similar to Mus musculus and Drosophila melanogaster nicastrin encoded by Genbank Accession Numbers AF240469 and AF240470; interacts with presenilin proteins PS1, PS2, S182, and E5-1"
	/codon_start=1
	/product="nicastrin"
	/protein_id="AAG11412.1"
	/db_xref="GI:9992878"
	/translation="MATAGGSGADPGSRGLRLISFCVLLAGLCRGSVERKIYIPLNTAPCVRLNATHQIGCOSSISGDTGVHIVKEEDLQWVLTGDPNPPYVLESKHETRDIMELKGRTRISRIAGLAVSLTKPSPASGFSPSVQPNDFGVYSNGPEFAHCR EIQWNSLNGLAYEDSFPIFLEDENETKIKQCYQDHLNLSONGSAFPLCAMOLF SHMAVISTATCMRRSSIOSTFENPEIVCDPLSDYNVMSMLKPIINTGLKPDPRV VATRDRSRFFWNVAGAEASAVASFTQLAAALOKAPDVTTLLPRNVAFVFEQGET FDYIGSSRMVYDMEKGFPEVLENDVDFELGOVALRSLIEMHTDPRVSOKESEVN OVEDLIATLEKSGAGVPAVILRRPNOSQPLPSSIQRLRLRNISGVVLADHSGAFHN KYOSIYDTAENINVSYPEWLSPEEDLNEFVDTAKALADAVATVLRALYELAGTNFS DTQVADPQTVTRLNLYFLIKANNSWFOILKQDRLSYLGDGPLQHYIAVSSPTNTYV VOYALANLTGTVYNLTREOCODPSKPSSENKDIYESWVGQPLHSNEDRLPRCVRSR ARLARALSPAFELSDMSSTSEYSTWTESRWMDIRARIFLIASKLELITLTVFGILIF SLIVTYCINAKADVFLFAPREPAVS"
BASE COUNT	682 a 803 c 692 g 770 t 2 others
ORIGIN	
Query Match	100.0%; Score 2948.2; DB 89; Length 2949;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2949; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 tctgcagaattcggcttgcgccttgaaacagaaacttcggctcttaggtccgggcca 60
Db	1 TCTGCAGAAATTCGGCTTGCGCTTGGAACACGAACTTCGGCTCTTAGCGCGGGCCA 60
QY	61 cagagacggtgtcagtgtagcctagagagcgctaacacagacagagcggaagggggc 120
Db	61 CAGAGACGCTGTAGTGTAGCCTAGAGAGCGCGCTAACAGACAGAGCCGAGCGGGGC 120
QY	121 ttccgctcagcagagagcgcaagatgctacgagcgaggggtgctctgggtcagccgg 180
Db	121 TTCCGCTCAGCAGAGAGGCAAGATGCGCTACGCGAGGGGTGGCTCTGGGGCTGACCCGGG 180
QY	181 aagtcggggtctccttcgcttcttctgtcgtcgtactagcaggttgtcagggg 240
Db	181 AAGTCGGGGTCTCTCTTGGCTTCTTCTTCTGCTCTACTAGCAGGTTGTGACAGGGG 240
QY	241 aaactcagtgagaggaagatatatcccccttaataaacaacagctccctgtgtcgct 300
Db	241 AAACCTCAGTGAGAGGAAGATATATCCCTTAATAAACAAGCTCCCTGTGTGCGCT 300
QY	301 gctcaacgcccactcatcagattgctgctccagctcttcaattagtgagacaggggtat 360
Db	301 GCTCAACGCCACTCATCAGATTGGCTGCCAGTCTTCAATTAGTGAGACACAGGGGTTAT 360
QY	361 ccaagtagtagaagaagaggaagcctacagtggtattagctatgagcccaaccccc 420
Db	361 CCACGTAGTAGAAGAAGAGGACCTACAGTGGGTATTGACTGATGATGCCCAACCCCC 420
QY	421 ttacatggttctgctgagagcaagcatttacaggagattaatgagaagctgaagg 480
Db	421 TTACATGTTCTGCTGAGAGCAAGCATTTTACACGAGGATTTAATGAGAAAGCTGAAGG 480
QY	481 gagaaccagccgaatgctgtctgacgtgtccttgaccaagccagtcctgcctcagg 540
Db	481 GAGAACCCAGCCGAATGCTGTCTTGACGTCTCTTGACCAAGCCAGTCCTGCTCAGG 540

QY	541	ctctctcctagtgtagcagtgccccaaatgtagtggttggtgttactccaattccataagg	600
Db	541	CTTCTCTCCTAGTGTACAGTGGCCCAAATGATGGGTTGGTGTACTCCAAATTCCTATGG	600
QY	601	gccagagtttgctcactgcagagaaatacagtgaattcgtctgggcaatggtttgctta	660
Db	601	GCCAGAGTTTGGTCTCACTGCAGAGAAATACAGTGGAAATTCGCTGGGCAATGGTTTGGCTTA	660
QY	661	tgaagaccttagttccccatcttctcttgaagatgaaatgaaaccaagtcatcaa	720
Db	661	TGAAGACCTTAGTTTCCCCATCTTCTTCTTGAGATGAAATGAAACCAAGTCATCAA	720
QY	721	gcagtgctatcaagatcacaaacctgagtgaagaatggtctcagcacaaccacttcccatatg	780
Db	721	GCAGTGTATCAGATCACAACTGAGTCAGAATGGCTCAGCACCAACTTCCCACTATG	780
QY	781	tgccatgcagctctttcacacatgcatgctgtlcatcagcactggccactgtcatgcygcg	840
Db	781	TGCCATGCAGCTTTTTCACATGCAATGCTGTTCATCAGCACTGCCACTGTGCATGCGGCG	840
QY	841	cagctccatccaagcaaccttcagcatccaaccagaatatcgltctgtaacccccgtctga	900
Db	841	CAGCTCCATCCAAAGCAACCTTCAGCATCAACCCAGAATCGTGTGTGACCCCTGTCTGA	900
QY	901	ttacaatgtgtgagcatgctaaagcctataatacaactgagacatlaaagcctgacga	960
Db	901	TTACAATGTGTGAGCATGCTAAAGCCTATAATAACAACCTGGACATTAAGCCTGACGA	960
QY	961	caggtgtgtgtgtgcgcacaccgcgctgtagtaagtcgtcccttttctggaatgtggcccc	1020
Db	961	CAGGTTGTGTGTGCTGCGCCACCCGCGTGAATAGTCTTCTTTCTGSAATGTGGCCCC	1020
QY	1021	aggggctgaaagcgcaagtgtgcttccctttgtcaoccagctgctgctgctggaagcttgca	1080
Db	1021	AGGGGCTGAAGCGCGCAGTGGCTTCTTGTCAACCACGCTGGCTGCTGTGAAGCTTTGCA	1080
QY	1081	aaaggcacctgtagtgcacacaccctgccccgcgaatgtcatgtttgtctcttccaagggga	1140
Db	1081	AAAGGCACCTGATGTGACCAACCTGCCCCGCAATGTCTATGTTGTCTTCAAGGGGA	1140
QY	1141	aacttttactacatctggcagctcgaagatggtctactacatatgagagaaggccaagttcc	1200
Db	1141	AACTTTTACTACTATTTGGCAGCTCGAGGATGGTCTACGATATGGAGAAGGGCAAGTTTCC	1200
QY	1201	cgtagcagttagagaatgttgaactcatctgtgagagctgggacaggtgcccctaagaacttc	1260
Db	1201	CGTGCAGTTAGAGAAATGTTGACTCATTTGTGAGAGCTGGGACAGGTGGCCTTAAGAACTTC	1260
QY	1261	attagagctttgtagtcacacacagatccgtttctcaagaaaaatgagtctgtaacgaacca	1320
Db	1261	ATTGAGAGCTTTGGATGCAACACAGATCCTGTTTCTCGAAAAAATGAGTGTGTACGGAAACA	1320
QY	1321	ggtgagagatctccctgcgcacacatlgagaagagtgctgctgtgctccctgctgcatcct	1380
Db	1321	GGTGAAGAGATCTCCCTGGCCACATTTGAGAAGAGTGGTGTGCTGTCCCTGCTCATCCT	1380
QY	1381	caggaagccaaatcagttcccaagcctctcccaaccatcttccctgcagcgatttctcgagc	1440
Db	1381	CAGGAGGCCAAATCAGTCCAGCCTCTCCACACCATCTTCCCTGCAGCGATTCTTCGAGC	1440
QY	1441	tcgnaacatctctgcggttgttctgtgctgaccactctgltgccttccataacaaataat	1500
Db	1441	TCGAAACATCTCTGGCGTGTCTTGCTGACCACTCTGTGCTTCATTAACAAATATTTA	1500
QY	1501	ccagagtatttaacgacactgctgagaaacattaatgttagactatcccgaatgctgagccc	1560
Db	1501	CCAGAGTATTTACGACACTGCTGAGAACATTAATGTGAGCTATCCCCGAATGGCTGAGCCC	1560
QY	1561	tgaagagagacctgaactttgtaacagacactgcgcaagcgccctggcagatgttggccaacggt	1620
Db	1561	TGAAGAGGACCTGAACCTTGTGTAAACAGACACTGCCAAGGCCCTGGCAGATGTGGCCACGCT	1620

QY	1621	gctgggaacgtgctctgtatagagcttgcaggaagaaaccaacttcagcgacacagttcaggc	1680
Db	1621	gctgggaacgtgctctgtatagagcttgcaggaagaaaccaacttcagcgacacagttcaggc	1680
QY	1681	tgatccccaacaggttacccgcctgctctatggttcctgatltaagccaacaactcatg	1740
Db	1681	tgatccccaacaggttacccgcctgctctatggttcctgatltaagccaacaactcatg	1740
QY	1741	gttccagtcctatcctcagcgaggaaactaagtlccctacttgggtgacgggccttcacaa	1800
Db	1741	gttccagtcctatcctcagcgaggaaactaagtlccctacttgggtgacgggccttcacaa	1800
QY	1801	ttacatcgctgtctccagcccccaacaaccactatgttgtacagtatgctcttgcca	1860
Db	1801	ttacatcgctgtctccagcccccaacaaccactatgttgtacagtatgctcttgcca	1860
QY	1861	tttgaactggcacagtggttcacacctcaccggagagcagtgccaagataaagtccc	1920
Db	1861	tttgaactggcacagtggttcacacctcaccggagagcagtgccaagataaagtccc	1920
QY	1921	aagtgaaaacaagatcctgtatgagtactcatggtgccagggcccttgcattcta	1980
Db	1921	aagtgaaaacaagatcctgtatgagtactcatggtgccagggcccttgcattcta	1980
QY	1981	gacggaaccgactcccccggtgtgtgcttctactgcacagattagccaggccttgtccc	2040
Db	1981	gacggaaccgactcccccggtgtgtgcttctactgcacagattagccaggccttgtccc	2040
QY	2041	tgccttctgaaactgagtcagtgagctctactgaatactctacatgtagtgagccgctg	2100
Db	2041	tgccttctgaaactgagtcagtgagctctactgaatactctacatgtagtgagccgctg	2100
QY	2101	gaaagataccggtgcccgatattctcatcgccagcaaaagagcttgagttgatcacct	2160
Db	2101	gaaagataccggtgcccgatattctcatcgccagcaaaagagcttgagttgatcacct	2160
QY	2161	gacagtgagcttcgcatcctcatctctccctcatcgtaacctactgcatcaatgccaa	2220
Db	2161	gacagtgagcttcgcatcctcatctctccctcatcgtaacctactgcatcaatgccaa	2220
QY	2221	agctgatgctcttctcatgtctccccgggagccaggaagctgtgcatactgagsagacc	2280
Db	2221	agctgatgctcttctcatgtctccccgggagccaggaagctgtgcatactgagsagacc	2280
QY	2281	scagcttctcttgccagctcagcagttcaacttcctagagcatctgcccactgggacaca	2340
Db	2281	scagcttctcttgccagctcagcagttcaacttcctagagcatctgcccactgggacaca	2340
QY	2341	accactaatgtgtcactcgtgaacctccctggcgctgtctcagattgggattaacataaa	2400
Db	2341	accactaatgtgtcactcgtgaacctccctggcgctgtctcagattgggattaacataaa	2400
QY	2401	agtgaacctatccaaaagagacaggaagaaataaataatgctccctcctccgctcc	2460
Db	2401	agtgaacctatccaaaagagacaggaagaaataaataatgctccctcctccgctcc	2460
QY	2461	ccttcccatcacccccctcccaattccctctcctctcactcaatgccaactaatgtgcc	2520
Db	2461	ccttcccatcacccccctcccaattccctctcctctcactcaatgccaactaatgtgcc	2520
QY	2521	ttacaataagaaagctctgtcctcgttttaactccctagttaccacacccctaattggc	2580
Db	2521	ttacaataagaaagctctgtcctcgttttaactccctagttaccacacccctaattggc	2580
QY	2581	tcaggaaccttctacttctccctccgcccgttacctctctcgtcctcaccacca	2640
Db	2581	tcaggaaccttctacttctccctccgcccgttacctctctcgtcctcaccacca	2640
QY	2641	cctgtaccagcacaactctcctgactgggaagagacataaaaggtttaatgtcaggytcaa	2700
Db	2641	cctgtaccagcacaactctcctgactgggaagagacataaaaggtttaatgtcaggytcaa	2700
QY	2701	ctacattgagccccctgaggaaggggcatctctgggctgagcctaagtctcctccac	2760



Db	2701	CTACATTGAGCCCGT	GAGGACAGGGGCATCTCTGGGCTGAGCC	TACTGTCTCTCCAC	2760
QY	2761	tgctcctttctccag	gccctcagatgycacat	taggtggtggtggtatcc	2820
Db	2761	TGTCTTTCTCCAGG	CCCTCAGATGGCACAT	TAGGGTGGGGCTGCGGGGTGGTATCC	2820
QY	2821	cacctccagccca	cagtgctcagttgtactttt	ttaagctgtaataatactat	2880
Db	2821	CACCTCCAGCCCA	CAGTGTCTGACTTTT	TAAAGCTGTAATATCTATTTTGT	2880
QY	2881	tttgtccttttcct	ttatctttt	gttaataataataataatgagtt	2940
Db	2881	TTTGTCTTTTCT	TTTATTTCTTTT	TGTAATATATAATGAGTTTCATTAATAATAGA	2940
QY	2941	ttatcccaac	2949		
Db	2941	TTATCCAC	2949		

RESULT	2		
D87442			
LOCUS	D87442	2805 bp	MRNA
DEFINITION	Human mRNA for KIAA0253 gene,		PRI
ACCESSION	D87442		10-JUL-1997
VERSION	D87442.1	GI:1665772	
KEYWORDS	KIAA0253.		
SOURCE	Homo sapiens male bone marrow myeloblast cell line:KG-1		cDNA to

## ORGANISM

REFERENCE 1 (bases 1 to 2805)

JOURNAL

REFERENCE	2 (sites)
AUTHORS	Seki, N., Ishikawa, K. and Nomura, N.
TITLE	Prediction of the coding sequences of unidentified human genes. VI.

JOURNAL  
REFERENCE  
AUTHORS

Unpublished (1996)  
3 (sites)  
Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayashi, Y.,

TITLE	LOCATION/QUALIFIERS
Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (K1AA0201-K1AA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain	
JOURNAL DNA Res. 3 (5), 321-329 (1996)	
MEDLINE 97191544	

```
FEATURES
source
location/Qualifiers
1. .2805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="KG-1"
/cell_type="myeloblast"
/clone_id="HA7036"
/tissue_type="PBLeucoscript II SK"
/sex="male"
/tissue_type="bone marrow"
1. .2128
/gene="KIAA0253"
<1. .2128
CDS
```

```

/gene="KIAA0253"
/codon_start=2
/protein_id="BAAL3383.1"
/db_xref="GI:1665773"
/translation="ATAGGSGADPGSRGILRLLSFVCLAGLCRGSVERKIYIPLN
KTAQVRLNATHQIGSSISGDTGVHVEKEEDLQWLTDGPNPVMYLLSKHF
TRDLMEKLKGRTSRIAGLAVSLTKPSPASGFSFVQCPNDGFGVYSNSYGEPAHCRE

```

BASE COUNT	648 a	765 c	646 g	746 t
ORIGIN	IQWNSLNGLAIVEDFSPIFILLEDENETKVIKQCYÖDHNLSONGSAPTFPLCAMÖLES HMHAVISTATCMRRSSIÖSTESINPEIYCDPLSDYNWWSMLKPIINTGTILKPDDRVV AATRLDSRSTFWNVNAPGASAVASEFTVÖLAAEALÖKAPDVTTLPNMEVFÖQETFE DYIGSSRMVYDMEKGKFPVÖLENENDFVELGÖVALRTSLELMHTDPVSÖKNEVRNÖ VEDLATLEKSGAVPAVILRRPNÖSÖPLBPSSÖLÖRFLARNISGVVLADHSGAFHNK YÖSITYDAENINVSYPEWLSPEDLNFTDTAKALADVATVLGRALYELAGGTNEFSD TVÖADPÖTVTRLLYGFLIKANNSWÖSILRÖDLRSYLGDPLOHYIAVSSPTNTTYV ÖYALANLTGTVNLJTRÖCÖDPSKVPSENÖDLEYSMVÖGPLHSHNETDRLRCVRSTA RLARALSPAFELÖWSSTEYSTWTESRKDIRARIFLIAKSELILITITVGFGLIFSS LITVYICNAKADVLFIAPREPGAVSY"			

Query Match	95.1%;	Score 2804.2;	DB 91;	Length 2805;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 2803; Conservative	2;	Mismatches	0;	Indels 0; Gaps 0;

QY	145	ggctacgycagggggtgycctctggygctgacccgggaagtgcgggtctccttcgcctct	204
Db	1	GGCTACGGCAGGGGGTGGCTCTGGGGGCTGACCCCGGAAGTCGGGGTCTCCTTCGCCCTTCT	60
QY	205	gtcttctcgcgtccctactagcaggtttgtgcagggggaaactcagtygagagaagatat	264
Db	61	GTCTTCTGCGTCCCTACTAGCAGGTTGTGTGCAGGGGAAACTCAGTGGAGAGGAAGATATA	120
QY	265	tatccccctaataaacaacgctccctgtgttcgcctgtcacaagccaactcatcagattg	324
Db	121	TATCCCCCTAAATATAAACAGCTCCCTGTGTTCGCCCTGCTCAACGCCACTCATCAGATTGG	180
QY	325	ctgccaagtcttcaattagttggagacacagggggtatccaagtagtagaagaagagga	384
Db	181	CTGCCAGTCTTCAATTAGTGGAGACACAGGGGTTATCCACGTAGTAGAAGAAAGAGGAGA	240
QY	385	cctacagtggtatattgactgatygcacccaacccccctlacatggtctgtgagagcaa	444
Db	241	CCTACAGTGGGTATTGACTGATGCCCCCAACCCCTTACATGTGTTCTGCTGGAGAGCAA	300
QY	445	gcattttaccagggattlaatgagagaagctgaaaaggagaaaccagccgaattgctgct	504
Db	301	GCAITTTACCAGGGAATTTAATGGAGAGACTGAAAGGGAGAACACGCCGAATTGCTGTCT	360
QY	505	tgcagtgctccttgacccaagcccagtcctgcctcaggtctctcctagtgtagacgtgcc	564
Db	361	TGCAGTGTCTTGACCAAGCCCACTGCTGCCTCAGGCTTCTCTCTAGTGTACAGTGCCC	420
QY	565	aaatgatgggttggtgttactccaattccctatggyccagagttgtcctcactgcagaga	624
Db	421	AAATGATGGGTTGTGTGTTTACTCCCAATTCCTATGGGCCAGAGTTTGTCTCACTGCAGAGA	480
QY	625	aatacagtggaattcgcctggyccaatggttggtctatgaaagcttagtttccccactt	684
Db	481	AATACAGTGAATTCGCTGGGCAATGTTGGCTTATGAAGACTTTAGTTTCCCACTCT	540
QY	685	tcttcttgaagatgaaaatgnaacccaagtcatcaagcagtgctatcaagatcacaacct	744
Db	541	TCTTCTTGAGATGAAAAATGAAACCAAGTCATCAAGCAGTGCATCAAGATCAACAACCT	600
QY	745	gagtcagaatggtctcagcaccacaccttcccactatgtgccaatgcagctctttcacacat	804
Db	601	GAGTCAGAATGGCTCAGACCAACCTTCCCACATATGTGCCATGCAGCTCTTTTCACACAT	660
QY	805	gcattgctgcatcagcaactgccaactgcatgcyggcgagctccatccaagaacaccttcag	864
Db	661	GCATGCTGTATCAGCACTGCCCACTGCATGGGGCCGACGCTCCATCCAAGCACCTTCAG	720
QY	865	catcaaccccaaaaatcgctctgtgaacccccctgtctgattacaatgtgtgagcatgtctaaa	924
Db	721	CATCAACCCAGAAATCGTCTGTGACCCCTGTGTGATTACAATGTGTGGAGCATGCTAAA	780
QY	925	gcctataatacaactggygacatlaaagcctgacgacagggltgtgtgtgctgccaaccg	984
Db	781	GCCTATAATAACAACCTGGGACATTTAAAGCCTGACGCACAGGGTTGTGTGTGCTGCCACCCG	840



QY 985 gctgtagtgcgtcccttcttctggaatgtgccccaggggctgaagaagcgcaagtgtctc 1044  
Db 841 GCTGGATAGTCTCTCTTTCTTGGAATGTGGCCCCAGGGGCTGAAGCGCAGTGGCTTC 900  
QY 1045 cttgtcaccagcgtgctgtcgtgaagccttgcaaaagcaccctgtatgacacccct 1104  
Db 901 CTTGTACACCCAGCTGGCTGCTGTGAAGCTTTGCAAAAGGCACTGATGTGACCACTC 960  
QY 1105 gccccgcaatgtcatgttcttcttcttcaagggaactttgactacatgtgacgtc 1164  
Db 961 GCCCCGCAATGTCTATGTTGCTCTTCTTCAAGGGGAAACTTTGACTACATTGGCAGCTC 1020  
QY 1165 gaagatgtctacagataggaaggaaggaagttcccggtgcaatgagaatgtgactc 1224  
Db 1021 GAGGATGCTCTACGATATGAGAGAGGCAAGTTCCCGTGACGTTAGAGATGTTGACTC 1080  
QY 1225 attgtgagcgtggaagaggtgaccttaagaacttcatagagccttgatgacacaga 1284  
Db 1081 ATTGTGAGCTGGGACAGGTGGCCCTTAAGAACTTCAATTAGAGCTTGTGATGACACAGA 1140  
QY 1285 tccgttctcagaaaaatgagctgtacaggaaccaggttgagagatctctgtgacatt 1344  
Db 1141 TCCTGTCTCTCAGAAAAATGAGTCTGTACGGAACAGGTGAGGATCTCTGGCCACATT 1200  
QY 1345 ggaagaagtggtgtgtgtccctgtgtcatctcaggaagggcccaatcagctccagcc 1404  
Db 1201 GGAGAAAGTGTGTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1260  
QY 1405 tctccacacatcttccctgagcgaatctcctcagagctcgaacaatctctgtgttct 1464  
Db 1261 TCTCCACCATCTTCCCTGACGAGATTCTTCTGAGCTGGAACATCTCTGTGTGTCT 1320  
QY 1465 ggtgacacactctgtgtcctcacaataatattaccagagatattaccagactgtga 1524  
Db 1321 GGCTGACCACTCTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1380  
QY 1525 gaacattaatgtagctataccgaatgtgagccctgaagaaggaacctgaactttgtaac 1584  
Db 1381 GAACATTAATGTAGCTATCCCGAATGTGCTGAGCCCTGAAGAGGACCTGAACCTTTGTAAC 1440  
QY 1585 agacactgcgaagggccctgtgagagatgtgccaaggtgtggaagctgtctgtatgagct 1644  
Db 1441 AGACACTGCCAAGGCCCTGGCAGATGTGGCCACGCTGTGGGACGTGTCTGTATGAGCT 1500  
QY 1645 tgcagaggaaccaacttccagcagacagttcaggtgatccccaacggttacccgct 1704  
Db 1501 TGCAGGAGGAACCAACTTCAGCGACAGTTCAGGCTGATCCCAAAACGGTTACCCGCT 1560  
QY 1705 gctctatggttctctgtatataagccaacactcatgttccagttcatctccagcagga 1764  
Db 1561 GCTCTATGGTCTCTGATTAAGGCCAACACTCATGTGTTCCAGTCTATCTCAGGCAAGGA 1620  
QY 1765 cctaaggtcctacttgggtgagcgggctcttcaacattacatcgtgtctccagcccccac 1824  
Db 1621 CCTAAGGTCCTACTTGGGTGACGGGCTCTTCAACATTACATCCGCTCTCAGACCCAC 1680  
QY 1825 caacaaccattatgttgtacagtatgcttggcaaatgtgactggcagcagtggtcaact 1884  
Db 1681 CAACACCACTTATGTTGTACAGTATGCTTGGCAAAATTGACTGTGGCAGAGTGGTCAACT 1740  
QY 1885 caaccgagagcagtgccaagatccaagtaaatcccaagtgaanaaagagatctgtatga 1944  
Db 1741 CACCCGAGAGCAGTGCAGAGATCCAAAGTAAGTCCCAAGTGAANAACAGGATCTGTATGA 1800  
QY 1945 gtactcatgtgtcagggcccttgcattcctaagagcaggaaccgactccccggtgtgt 2004  
Db 1801 GTACTCATGGGTCCAGGGCCCTTGTGATTTATGAGACGGACCGACTCCCGGTGTGT 1860  
QY 2005 gcgttactgacagattagccagggcctgtctcctgaccttgaactgaagtcagtgag 2064  
Db 1861 GCGTCTACTGACAGATTAAGTACGAGGCGCTTGTCTCTGCTTGTGAACGTAGTCACTGAG 1920

QY 2065 ctctactgaactctacatgagactgagagccgctggaagaatatccggtcccgatatt 2124  
Db 1921 CTCTACTGAACTCTACATGAGCTGAGAGCCCGCTGGAAAGATATCCGTGGCCGATATT 1980  
QY 2125 tctcatcgccagcaaaagagctgtgagttgatcaaccctgacagtggtcctggcatctcat 2184  
Db 1981 TCTCATCGCCAGCAAAAGAGCTGAGTGTGATCACCCTGACAGTGGGCTTCGCACTCAT 2040  
QY 2185 ctctccctcatcgctcacctactgcatcaaatgccaagctgatgtccttctcatgtctc 2244  
Db 2041 CTCTCCCTCATGCTCACCCTACTGATCAATGCCAAAGCTGATGCTCTTTCATTTGCTCC 2100  
QY 2245 ccggaagccagagctgtgtcatactgagagagaccsagccttcttgccagctcagca 2304  
Db 2101 CCGGAGCCAGAGCTGTGTATCTAGAGAGACCCCAAGCTTTCTTGGCAGCTCAGCA 2160  
QY 2305 gttcacttctcagacatctgtcccaactgggacacaaccaatattgtcactggaacct 2364  
Db 2161 GTTCACTTCTTAGCATCTGTCCCACTGGGACACAAACCACTAATTGTCACTGGAACCT 2220  
QY 2365 ccctggcctgtctcagatttggaatcaataaagagtggaactatccaagaagacag 2424  
Db 2221 CCCTGGGCTGTCTCAGATTGGGATTAACATAAAAGAGTGAACATATCCAAAAGAGACAG 2280  
QY 2425 ggaagaataaataatgtccctccctccctccctccctccctccctccctccctccct 2484  
Db 2281 GGAGAAATTAATTAATTTGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2340  
QY 2485 tctctctctctctctactcatgacagatattgggattacaataagaagctctgtccc 2544  
Db 2341 TTCCTCTCTCTCTCTACTCATGCGAGATTGGGATTACAATAAGAAAGCTTCTGTGCTCC 2400  
QY 2545 tgtttaactccctagttaccacaccccaatattggccctcaggaacctctacttctcct 2604  
Db 2401 TGTTTAACTCCCTAGTTACCCACCCCTAATTTGGCCCTTCAAGACCCCTTCTACTTTTCT 2460  
QY 2605 cctgcccctgtacctctctctgtctcctcaacccccccccctgtaccagccaactctgac 2664  
Db 2461 CCTGCCCTGTACCTCTCTCTGTCTCTCTCAACCCCAACCCCTGTACCCGACCACTTCTGTAC 2520  
QY 2665 tgggaagacataaaagtttaattgtcagaggtgtcaaaactacattgagccctgagagag 2724  
Db 2521 TGGGAAGACATAAAAGTTTAATGTCAAGGTTCAAACTACATTGAGGCCCTGAGGACAGG 2580  
QY 2725 ggcattctgtgctgagcctactgtctcctcccaactgtccttctccagggccctcagat 2784  
Db 2581 GGCATCTCTGGGCTGAGCCTACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2640  
QY 2785 ggcacataaggtgtggt 2844  
Db 2641 GGCACATTAAGGCTGGGCTGTGGGGGTGGGTATCCACCTCCAGGCCACAGTGTCAAGT 2700  
QY 2845 tgtacttttataagctgtataatctattttgttttgtcttcttcttcttcttcttctt 2904  
Db 2701 TGTACTTTTATAAGCTGTAATATCTATTTTGTGTTTGTCTTTTCTTTTCTTTTCTTTT 2760  
QY 2905 tgtaaatatatatataatgagttcattaaatatagattatccac 2949  
Db 2761 TGTAAATATATATATATATAGATTCAATTAATAATAGATTATCCAC 2805

RESULT 3  
AF240469 2250 bp mRNA ROD 08-SEP-2000  
LOCUS AF240469  
DEFINITION Mus musculus nicastrin mRNA, complete cds.  
ACCESSION AF240469  
VERSION AF240469.1 GI:9992879  
KEYWORDS house mouse.  
SOURCE Mus musculus.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2250)



Db 1318 ACATCTTGCGCTGGTCTGGCTGACCACTCTGGCTCCCTCCACATCGTATACCAGA 1377  
QY 1506 gtaattacgacactgctgagaacattaatgtgactatcccgaaatgctgagccctgaag 1565  
Db 1378 GCATTATGACACCGGCTGAGAACATTAATGTGACCTATCCCTGAGTGGCAGAGCCCAAG 1437  
QY 1566 agaacctgaacttgttaacagacactgccaaggccctgagcagatgtgcccaggtgtg 1625  
Db 1438 AGGACCTCAACTTGTGACAGACACTGCCAAGGCACTGGCGCAATGTGCCACAGTGTGG 1497  
QY 1626 gacgtgctgtatgagctgtgcaaggaaaccaacttcagcgacacagttcaggtgatc 1685  
Db 1498 CGCGTGACCTGATGAGCTTGACAGGAGAACCACTTCAGCAGCTCCATCCAGGCTGATC 1557  
QY 1686 cccaacggttacccgctgctctatggttcctgattaaagccaacaactcatgttcc 1745  
Db 1558 CCCAGACAGTTACAGCTCTGCTTATGGGTTCTGTTAAAGCTAACACTCATGTTTC 1617  
QY 1746 agtctatcctcagcagagaccctaaggtcctacttgggtgagcggccctcacaattaca 1805  
Db 1618 AGTCGATCCTGAACATGACCTAAGGTCCTATTGATGACAGGCGCTTCAACTACA 1677  
QY 1806 tcgctgtctccagccccaacacacacactatgtgtgtacagtatgccttggcacaattga 1865  
Db 1678 TCGCCGTTCCAGCCCTACCAACAGACTTACGTTGTGACAGTACGCTTGCCAAACCTGA 1737  
QY 1866 ctggcacagtggtcaacctcaccgagagagcaggtgccaagtccaagtccaagt 1925  
Db 1738 CTGGCAAGCGGACCAACTCACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797  
QY 1926 aaacaagatctgtatgagtagtactatggtgtccaggcccttgcattcctaagagag 1985  
Db 1798 AGAGCAAGGATTTATATGATACTCGTGGTACAGAGCCCTTGGAATCCAAAGAGAGAG 1857  
QY 1986 accgactcccccggtgtgtggttctactgcagatagcagggcctgtgtcctgct 2045  
Db 1858 AGAGGCTCCACAGTGTGTGGCTCCACAGTGGAGTGGCCAGGCGCTTGCCCTGCT 1917  
QY 2046 ttgaactgagtcagtgaggtctactatgaatactctacatgagtagagccgtggaag 2105  
Db 1918 TTGAACGTGAGTCAAGTGGAGCTCCACAGATACTCTACGTGGGCGGAGAGCGCGTGAAG 1977  
QY 2106 atatcgtgccccgataatcttcacatgcgcagcaagagcttgatgataccctgacag 2165  
Db 1978 ACATCCAAGCTCGGATATTCCTAATTGCCAGCAAAAGCTTGAGTTCAACGCTGATCG 2037  
QY 2166 tggcctcgcacatcctcatctctccctcatcgtcaacctactgcatcaatgccaaagc 2225  
Db 2038 TGGGCTTCAGCATCTTATCTTCTCTCATCTGTCACCTACTGCAATGCCAAAGCG 2097  
QY 2226 atgtccttcatgtgtccccggagcagagagctgtgcatcactgagsagaccscagc 2285  
Db 2098 ACGTCCCTTGTGTGCTCCCGAGAGGAGGAGGCTGTGTCTTACTGAAGAGGACTCTAGC 2157  
QY 2286 ttctctgcaagctcagcaggttcaactcctagagcatgttccactgagcacaaccac 2345  
Db 2158 TCTCCCTGCTGCTCTGAACCTTACTTCCAGACAGAGGTGTCCGGCTGGGAACAAACAC 2217  
QY 2346 taattgtcactggaacctccctggtgctg 2375  
Db 2218 TAATTTGTACAGTGTCTCTGGGCTG 2247

RESULT 4  
AL513282/c AL513282 155913 bp DNA HTG 26-FEB-2001  
LOCUS Homo sapiens chromosome 1 clone RP11-179G5, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\* 8 unordered pieces.

ACCESSION AL513282  
VERSION AL513282.5 GI:13162044  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 155913)  
Pavitt, R.  
Direct Submission  
Submitted (25-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Feb 28, 2001 this sequence version replaced g1:13161719.  
----- Genome Center

COMMENT

Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: BA179G5  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 153015 bases at least Q40  
Consensus quality: 153938 bases at least Q30  
Consensus quality: 154565 bases at least Q20  
Insert size: 155213; sum-of-contigs  
Insert size: 164818; 1.0% error; agarose-fp  
Quality coverage: 6.89x in Q20 bases; sum-of-contigs Quality  
coverage: 7.27x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 20838: contig of 20838 bp in length  
\* 20839 20938: gap of 100 bp  
\* 20939 50589: contig of 29651 bp in length  
\* 50590 50689: gap of 100 bp  
\* 50690 81538: contig of 30849 bp in length  
\* 81539 81638: gap of 100 bp  
\* 81639 89682: contig of 8044 bp in length  
\* 89683 89782: gap of 100 bp  
\* 89783 108105: contig of 18323 bp in length  
\* 108106 108205: gap of 100 bp  
\* 108206 149395: contig of 41190 bp in length  
\* 149396 149495: gap of 100 bp  
\* 149496 153075: contig of 3580 bp in length  
\* 153076 153175: gap of 100 bp  
\* 153176 155913: contig of 2738 bp in length.

FEATURES  
source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

1. 155913  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-179G5"  
/clone\_lib="RPC1-11.1"  
1. 20838  
/note="assembly\_fragment:02743  
fragment\_chain:1  
clone\_end:T7  
vector\_side:left"  
20939. 50589  
/note="assembly\_fragment:01236  
fragment\_chain:1"  
50690. 81538  
/note="assembly\_fragment:00478  
fragment\_chain:2"  
81639. 89682  
/note="assembly\_fragment:03283  
fragment\_chain:2"  
89783. 108105



```

misc_feature      /note="assembly_fragment:00942
                  fragment_chain:3"
                  108206. 149395
misc_feature      /note="assembly_fragment:02750
                  fragment_chain:3"
                  149496. 153075
misc_feature      /note="assembly_fragment:02486
                  fragment_chain:4"
                  153176. 155913
misc_feature      /note="assembly_fragment:03083
                  fragment_chain:4"
                  34529  c  35326  g  41527  t
BASE COUNT      43830  a
ORIGIN

```

Query Match	27.28;	Score 802;	DB 81;	Length 155913;
Best Local Similarity	99.58;	Pred. No. 2.6e-204;		
Matches 802; Conservative	2;	Mismatches 2;	Indels 0;	Gaps 0;

QY	2144	ctgagttgatcaacccctgaacagltgggctcgacatccatcttccctcatcgtcac	2203
Db	95606	CTGCAGTGTGATCACCCCTGACAGTGGGCTTCGGATCCTCATCTTCCTCATCGTCACC	95547
QY	2204	tactgcatacatgcccacaaagctgatgtccttctcatlgtccccggagccagagctgtg	2263
Db	95546	TACTGCATCAATGCCAAGCTGATGTCTTTTCATTGTGCTCCCGGGAGCCAGAGAGCTGTG	95487
QY	2264	tcatactgagssgggaccscagcttltcttgccaagctcagcagttcaacttccctagacatc	2323
Db	95486	TCATTA CTGAGGAGGAGACCCCAAGCTTTTCTTGCCAGCTCAGCAGTTCACTTCTTAGAGCATC	95427
QY	2324	tgtcccaactggggacacacacccaactaatlgtgaactygaacctcccctgggacctgtcagat	2383
Db	95426	TGTCCCACTGGGACACACCACTAATTGTGTCACTGGAACCTCCCTGGGCCGTGTCTCAGAT	95367
QY	2384	tyggattaacataaaagaagltggaactatcccaaaaagagacagggagaaataaataatgc	2443
Db	95366	TGGGATTAACATAAAAGAGTGGAACTATCCAAAAAGAGACAGGAGAGAAATAAATAATTGC	95307
QY	2444	ctcccttccctccgctccccccttcccaatcacccctccccatctctctctctact	2503
Db	95306	CTCCCTTCTCTCCGCTCCCTTCCCATCAACCCCTTCCCAATTCTCTTCTCTCTCTACT	95247
QY	2504	catgcagagatlltgggattacaataagaagctctgtctcctgtttaactccctagttac	2563
Db	95246	CATGCCAGATTTTGGGATTACAAATAGAAAGCTTCTGTGCTCTGTTAACTCCCTAGTTAC	95187
QY	2564	ccaccctaatttggccctcagaagcccttctaacttltccctcctgcccgtacactctc	2623
Db	95186	CCACCCTAATTTGGCCCTTCAGGACCCCTTCTACTTTTTCCTTCTCTGCCCCGTACTCTCTC	95127
QY	2624	tgtctcctcaaccccccaacccctgttaaccagaaccttcttgactyggaaagacataaaagt	2683
Db	95126	TGCTTCTCACCCCAACCCCTGTATCCAGCCACCTTCTGACTGGGAAGACATATAAAGT	95067
QY	2684	ttaatgtcaggytcaaacatacatlgagccctgagagacagggcatctctgygctgagcc	2743
Db	95066	TTAATGTCAGGGTCAAACTACATTGAGCCCCCTGAGGACAGGGGCATCTCTGGGCTGAGCC	95007
QY	2744	tactgtctccttcccactgtccttltctccagggccctcagatgycacattaggytgycgt	2803
Db	95006	TACTGTCTCTTCCCACTGTCTTTCTCCAGGGCCCTCAGATGGCACATTAAGGTGGGCGT	94947
QY	2804	gctgcggytgggtatcccaacctccagcccacagltgctcagltgtacttttataaagctg	2863
Db	94946	GCTGCGGGTGGGTATPCCACCTCCAGGCCACAGTGTGCTACTTTTATTAAGCTG	94887
QY	2864	taatatcatttttgtttttgtcttlttcccttattcttlttgttaatatataataatg	2923
Db	94886	TAATATCTATTTTGTGTTTGTCTTTTCCCTTAATCTTTTGTAAATATATATAATATG	94827
QY	2924	agtttcaatlaaanaagattatccac	2949

Db 94826 AGTTTCATTAAATAGATTATCCAC 94801

RESULT	5
AL445230	
LOCUS	AL445230 182344 bp DNA HTG 10-MAR-2001
DEFINITION	Homo sapiens chromosome 1 clone RP11-517F10, *** SEQUENCING IN PROGRESS ***, 13 unordered pieces.
ACCESSION	AL445230
VERSION	AL445230.4 GI:13277499
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 182344)
AUTHORS	McLay, K.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB0 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clome requests: clomerequest@sanger.ac.uk
COMMENT	On Mar12, 2001 this sequence version replaced at:13274828

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA517F10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 176921 bases at least Q40
Consensus quality: 178814 bases at least Q30
Consensus quality: 179919 bases at least Q20
Insert size: 181144; sum-of-contigs
Insert size: 187578; agarose-fp
Quality coverage: 5.35x in Q20 bases; sum-of-contigs Quality
coverage: 5.28x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 6697: contig of 6697 bp in length
* 6698 6797: gap of 100 bp
* 6798 42129: contig of 35332 bp in length
* 42130 42229: gap of 100 bp
* 42230 54864: contig of 12635 bp in length
* 54865 54964: gap of 100 bp
* 54965 62235: contig of 7271 bp in length
* 62236 62335: gap of 100 bp
* 62336 81626: contig of 19291 bp in length
* 81627 81726: gap of 100 bp
* 81727 88611: contig of 6885 bp in length
* 88612 88711: gap of 100 bp
* 88712 95096: contig of 6385 bp in length
* 95097 95196: gap of 100 bp
* 95197 97751: contig of 2555 bp in length
* 97752 97851: gap of 100 bp
* 97852 107298: contig of 9447 bp in length
* 107299 107398: gap of 100 bp
* 107399 136764: contig of 29366 bp in length
* 136765 136864: gap of 100 bp
* 136865 140326: contig of 3462 bp in length
* 140327 140427: gap of 101 bp
* 140428 164080: contig of 23653 bp in length
* 164081 164180: gap of 100 bp

```

```

FEATURES
  * 164181 182344: contig of 18164 bp in length.
  source
    1. 182344
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="1"
    /clone="RP11-517F10"
    /clone_11b="RP11-11.2"
    1. 6697
  misc_feature
    /note="assembly_fragment:00222"
    fragment_chain:1
    clone_end:SP6
    vector_side:left"
    6798. 42129
  misc_feature
    /note="assembly_fragment:02218"
    fragment_chain:1"
    42230. 54864
  misc_feature
    /note="assembly_fragment:00307"
    fragment_chain:1"
    54965. 62235
  misc_feature
    /note="assembly_fragment:02611"
    fragment_chain:1"
    62336. 81626
  misc_feature
    /note="assembly_fragment:02239"
    fragment_chain:2"
    81727. 88611
  misc_feature
    /note="assembly_fragment:01749"
    fragment_chain:2"
    88712. 95096
  misc_feature
    /note="assembly_fragment:01512"
    fragment_chain:2"
    95197. 97751
  misc_feature
    /note="assembly_fragment:01990"
    fragment_chain:2"
    97852. 107298
  misc_feature
    /note="assembly_fragment:00285"
    fragment_chain:3"
    107399. 136764
  misc_feature
    /note="assembly_fragment:00288"
    fragment_chain:3"
    136865. 140326
  misc_feature
    /note="assembly_fragment:02710"
    fragment_chain:3"
    140427. 164080
  misc_feature
    /note="assembly_fragment:01259.0"
    164181. 182344
  misc_feature
    /note="assembly_fragment:01700"
    53009 a 40249 c 38908 g 48952 t 1226 others
  ORIGIN
  Query Match 27.2%; Score 802; DB 81; Length 182344;
  Best Local Similarity 99.5%; Pred. No. 2.6e-204;
  Matches 802; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
  QY 2144 ctgaggtgacacccgacagtggtggtcgagcatcctcattctccctcatcgacac 2203
  11 |||||
  Db 19292 CTGCAGTGTATCACCTGACAGTGGGCTTCGGCATCTCTCATCTTCTCCCTCATCGTCACC 19351
  |||||
  QY 2204 tactgcatcaatgccaagctgatgtccttttcatctgtctcccgaggagccagagctgtg 2263
  |||||
  Db 19352 TACTGCAATCAATGCCAAGCTGATGCTTTTCATTTGCTCTCCCGGAGCCAGAGCTGTG 19411
  |||||
  QY 2264 tcatactgagsagaccagcttcttctgcccagctcagcagttcattcctcctagagcatc 2323
  |||||
  Db 19412 TCATAGTGAAGAGAGACCCAGCTTTCTTGGCCAGCTCAGCAGTTCACCTTCTAGAGCATC 19471
  |||||
  QY 2324 tgtccactgagacacacacacacacacacacacacacacacacacacacacacacacacac 2383
  |||||
  Db 23472 TGTCCGACTGGGACACACACACACACACACACACACACACACACACACACACACACAC 19531
  |||||
  QY 2384 tgggattaacataaagagtggaactatccaaagagacagggagaataaataatgc 2443
  |||||

```

```

Db 19532 TGGGATTACATATAAGAGTGAACATATCCAAAAGAGACAGGAGAAATAAATTGC 19591
QY 2444 ctccttcctcgcgtcccttcctccatcacccctccctcattctcctcctact 2503
|||
Db 19592 CTCCCTTCTCCGCTCCCTTTCCCATCACCCCTTCCCATTTCTCTCTCTACT 19651
|||
QY 2504 catgccaagatttgggatacaatagaagcttctgtcctcgtttaactccctagttac 2563
|||
Db 19652 CATGCCAGATTTGGATACAAATAGAAGCTTCTGTCTCCTGTTAACTCCCTAGTTAC 19711
|||
QY 2564 ccaccctaatgtgcccctcagagccctcctacttttctcctcctcgtacccctc 2623
|||
Db 19712 CCACCTAATTTGCCCTTCAGAGACCCCTTCTACTTTTCTCTCTCCCTGACCTCTC 19771
|||
QY 2624 tgtcctcaccaccaccctcgtacccagccacccctcctcgtactggaagacataaaggt 2683
|||
Db 19772 TGCTCCTACACCCACCCCTGTACCCAGCCACCTTCTGACTGGAAAGACATAAAGGT 19831
|||
QY 2684 ttaatgcaaggtcaaaactacatgaagccctcagagcaagggagcctctggctgagcc 2743
|||
Db 19832 TTAATGTACAGGTCAAACTACATTGAGCCCTTGAGGACAGGGGCATCTCTGGCTGAGCC 19891
|||
QY 2744 tactgtccttcctccactgtccttctcagagccctcagatgacacattaggtggcgt 2803
|||
Db 19892 TACTGTCTCTTCCCATCTCTCTTCTCTCAGGCCCTCAGATGGCACATTAGGGTGGCGCT 19951
|||
QY 2804 gctgcgggtgggtatccacccctcagccacccagcagtgctcagttgacttttataagctg 2863
|||
Db 19952 GCTGCGGTGGGTATCCCATCTCCAGCCACAGCTGCTCAGTGTACTTTTATTAAGCTG 20011
|||
QY 2864 taatatcattttgttttgccttttctccttattcttlttgataatatataatg 2923
|||
Db 20012 TAATATCTAATTTGTTTGTCTTTTCTTTCTTTATTTGTAATATATATATATG 20071
|||
QY 2924 agttcattaaatagattatccac 2949
|||
Db 20072 AGTTTCATTAATAATAGATTATCCAC 20097
|||

```

```

RESULT 6
AL138932 170677 bp DNA HTG 14-JUN-2000
LOCUS
DEFINITION
  Homo sapiens chromosome 1 clone RP1-24A65 map p11.1-13.3, ***
  SEQUENCING IN PROGRESS ***, 31 unordered pieces.
ACCESSION
  AL138932
  AL138932.10 GI:9230945
  HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
  SOURCE
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 170677)
  Pavitt,R.
  Direct Submission
  Submitted (13-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
  requests: clonerequest@sanger.ac.uk
  On Jul 16, 2000 this sequence version replaced gi:8573815.
  ----- Genome Center
  Center: Sanger Centre
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquerry@sanger.ac.uk
  Project Information
  ----- Summary Statistics
  Center project name: dj24465
  Assembly program: XGAP4; version 4.5
  Sequencing vector: plasmid; I08752; 100% of reads
  Chemistry: Dye-terminator Big Dye; 39% of reads
  Chemistry: Dye-terminator ET-amersham; 60% of reads
  Consensus quality: 156012 bases at least Q40
  Consensus quality: 160965 bases at least Q30
  Consensus quality: 164137 bases at least Q20

```

```

COMMENT
  Center: Sanger Centre
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquerry@sanger.ac.uk
  Project Information
  ----- Summary Statistics
  Center project name: dj24465
  Assembly program: XGAP4; version 4.5
  Sequencing vector: plasmid; I08752; 100% of reads
  Chemistry: Dye-terminator Big Dye; 39% of reads
  Chemistry: Dye-terminator ET-amersham; 60% of reads
  Consensus quality: 156012 bases at least Q40
  Consensus quality: 160965 bases at least Q30
  Consensus quality: 164137 bases at least Q20

```

Insert size: 167677; sum-of-contigs  
 Insert size: 131821; 35.2% error; agarose-fp  
 Quality coverage: 3.70x in Q20 bases; sum-of-contigs Quality  
 coverage: 4.70x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently  
 consists of 31 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 4821: contig of 4821 bp in length  
 4822 4921: gap of 100 bp  
 4922 11645: contig of 6724 bp in length  
 11646 11745: gap of 100 bp  
 11746 13929: contig of 2184 bp in length  
 13930 14029: gap of 100 bp  
 14030 16875: contig of 2846 bp in length  
 16876 16975: gap of 100 bp  
 16976 20388: contig of 3413 bp in length  
 20389 20489: gap of 100 bp  
 20489 29256: contig of 8768 bp in length  
 29257 29356: gap of 100 bp  
 29357 30470: contig of 1114 bp in length  
 30471 30570: gap of 100 bp  
 30571 34082: contig of 3512 bp in length  
 34083 34182: gap of 100 bp  
 34183 49557: contig of 15375 bp in length  
 49558 49657: gap of 100 bp  
 49658 51777: contig of 2120 bp in length  
 51778 51877: gap of 100 bp  
 51878 53122: contig of 1245 bp in length  
 53123 53222: gap of 100 bp  
 53223 56791: contig of 3569 bp in length  
 56792 56891: gap of 100 bp  
 56892 61955: contig of 5064 bp in length  
 61956 62055: gap of 100 bp  
 62056 70347: contig of 8292 bp in length  
 70348 70447: gap of 100 bp  
 70448 77556: contig of 7109 bp in length  
 77557 77656: gap of 100 bp  
 77657 81110: contig of 3454 bp in length  
 81111 81210: gap of 100 bp  
 81211 82231: contig of 1021 bp in length  
 82232 82331: gap of 100 bp  
 82332 91668: contig of 9337 bp in length  
 91669 91768: gap of 100 bp  
 91769 112272: contig of 20504 bp in length  
 112273 112372: gap of 100 bp  
 112373 120545: contig of 8173 bp in length  
 120546 120645: gap of 100 bp  
 120646 140354: contig of 19709 bp in length  
 140355 140454: gap of 100 bp  
 140455 141779: contig of 1325 bp in length  
 141780 141879: gap of 100 bp  
 141880 143172: contig of 1293 bp in length  
 143173 143272: gap of 100 bp  
 143273 150964: contig of 7692 bp in length  
 150965 151064: gap of 100 bp  
 151065 152556: contig of 1492 bp in length  
 152557 152656: gap of 100 bp  
 152657 153707: contig of 1051 bp in length  
 153708 153807: gap of 100 bp  
 153808 154871: contig of 1064 bp in length  
 154872 154971: gap of 100 bp  
 154972 156380: contig of 1409 bp in length  
 156381 156480: gap of 100 bp  
 156481 160084: contig of 3604 bp in length  
 160085 160184: gap of 100 bp  
 160185 163849: contig of 3665 bp in length  
 163850 163949: gap of 100 bp

FEATURES	*	163950	170677: contig of 6728 bp in length.
source		1. .170677	location/Qualifiers
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="1"	
		/map="p11.1-13.3"	
		/clone="RP1-244G5"	
		/clone_lib="RPC1-1"	
		1. .4821	
misc_feature		/note="assembly_fragment:01410"	
		fragment_chain:1"	
misc_feature		4922. .11645	
		/note="assembly_fragment:00944"	
		fragment_chain:1"	
misc_feature		11746. .13929	
		/note="assembly_fragment:00895"	
		fragment_chain:1"	
misc_feature		14030. .16875	
		/note="assembly_fragment:01274"	
		fragment_chain:1"	
misc_feature		16976. .20388	
		/note="assembly_fragment:00718"	
		fragment_chain:1"	
misc_feature		20489. .29256	
		/note="assembly_fragment:00242"	
		fragment_chain:1"	
misc_feature		29357. .30470	
		/note="assembly_fragment:00201"	
		fragment_chain:1"	
misc_feature		30571. .34082	
		/note="assembly_fragment:01345"	
		fragment_chain:1"	
misc_feature		34183. .49557	
		/note="assembly_fragment:01469"	
		fragment_chain:1"	
misc_feature		49658. .51777	
		/note="assembly_fragment:01230"	
		fragment_chain:2"	
misc_feature		51878. .53122	
		/note="assembly_fragment:00199"	
		fragment_chain:2"	
misc_feature		53223. .56791	
		/note="assembly_fragment:01418"	
		fragment_chain:2"	
misc_feature		56892. .61955	
		/note="assembly_fragment:01391"	
		fragment_chain:2"	
misc_feature		62056. .70347	
		/note="assembly_fragment:01376"	
		fragment_chain:3"	
misc_feature		70448. .77556	
		/note="assembly_fragment:00605"	
		fragment_chain:3"	
misc_feature		77657. .81110	
		/note="assembly_fragment:00470"	
		fragment_chain:3"	
misc_feature		81211. .82231	
		/note="assembly_fragment:00573"	
		fragment_chain:3"	
misc_feature		82332. .91668	
		/note="assembly_fragment:01058"	
		fragment_chain:4"	
misc_feature		91769. .112272	
		/note="assembly_fragment:00207"	
		fragment_chain:4"	
misc_feature		112373. .120545	
		/note="assembly_fragment:01906"	
		fragment_chain:4"	
misc_feature		120646. .140354	
		/note="assembly_fragment:00800"	
		fragment_chain:5"	
misc_feature		140455. .141779	





TITLE	JOURNAL	COMMENT
Direct Submission	Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)	On May 30, 2000 this sequence version replaced gi:7717300. The Chromosome 21 mapping and sequencing consortium consisting of * RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagami-hara 228-8555, Japan, * e-mail: hattori@gsc.riken.go.jp * URL: http://hgp.gsc.riken.go.jp/ and
	* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany, * e-mail: gscj-submit@genome.imb-jena.de * URL: http://genome.imb-jena.de/ and	
	* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan, * e-mail: nshimizu@dmf.med.keio.ac.jp * URL: http://www.dmf.med.keio.ac.jp/ and	
	* GBF, Dept. of Genome Analysis, * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e-mail: info.genome@gbf.de * URL: http://genome.gbf.de/ and	
FEATURES	* Max-Planck Institute for Molecular Genetics, * Ihnestrasse 73, D-14195 Berlin, Germany, * e-mail: info-chr21@molgen.mpg.de * URL: http://chr21.tz-berlin.mpg.de/ AL163244: Submitted (10-Apr-2000). Location/Qualifiers	
source	1.340000	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="21"	
	<1.159143	
	/note="Accession No. AP001604"	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="21"	
	/map="21q21.2"	
	/clone="KB164888, 5' partial"	
	/clone_lib="Keio BAC library"	
	4113.4336	
	/note="L2"	
repeat_region	/rpt_family="LINE/L2"	
	/rpt_type=DISPERSED	
	5431.5736	
	/note="L1ME3A"	
repeat_region	/rpt_family="LINE/L1"	
	/rpt_type=DISPERSED	
	complement(5970..6279)	
	/note="AluO"	
repeat_region	/rpt_family="SINE/Alu"	
	/rpt_type=DISPERSED	
	complement(6306..6657)	
	/note="THE1B"	
repeat_region	/rpt_family="LTR/MaLR"	
	/rpt_type=DISPERSED	
	7231..7524	
	/note="LTR16A"	
repeat_region	/rpt_family="LTR/Retroviral"	
	/rpt_type=DISPERSED	
	complement(7526..7906)	
	/note="THE1C"	
repeat_region	/rpt_family="LTR/MaLR"	

repeat_region	/note="HAL1" /rpt_family="LINE/Other" /rpt_type=DISPERSED complement(8526. .8895) /note="LIM4c" /rpt_family="LINE/L1" /rpt_type=DISPERSED 8896. .9254 /note="THE1B" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED complement(9255. .9447) /note="LIM4c" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(9496. .9676) /note="AluJb" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(9769. .10321) /note="LIM4c" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(10880. .11274) /note="LIMD" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(12716. .13021) /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 14710. .14995 /note="AluY" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(15316. .15687) /note="MER65-internal" /rpt_family="LTR/MER4_group" /rpt_type=DISPERSED 15689. .15860 /note="MLT1A1" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED 15949. .16335 /note="MLT1J2" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED 17821. .18128 /note="AluSg" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 18313. .18417 /note="MER63" /rpt_family="DNA/MER1_type" /rpt_type=DISPERSED complement(20658. .20797) /note="L1PA4" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(21855. .23050) /note="L1PA16" /rpt_family="LINE/L1" /rpt_type=DISPERSED 23051. .23409 /note="THE1B" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED complement(23410. .24907) /note="L1PA16" /rpt_family="LINE/L1" /rpt_type=DISPERSED 24908. .24927 /note="(TTC)n"
---------------	---

[illegible]



QY 2831 ccacagtcgtcagttgtac-ttttataagctgtaataatctattttttgtcttt 2889  
 Db 205023 CTTTCAGTGTCTCAGTGTACTTTTATTAAGCTATAATATCTATTTT----- 205069  
 QY 2890 ttcttatttctttttgtataataatgagtttcaataaatagattatccc 2947  
 Db 205070 -----TTTCTTTTTTTGTAAATATTAATAATATTGAGTTTCATTAATAATAGATAACCC 205123

RESULT 9  
 AC090058/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome 12q clone RP11-112N23, \*\*\* SEQUENCING IN  
 AC090058  
 VERSION AC090058  
 KEYWORDS AC090058.6 GI:13194919  
 SOURCE HTG: HTGS\_PHASE1.  
 ORGANISM human.  
 Homo sapiens

REFERENCE  
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,  
 Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,  
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,  
 Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
 Kovar,C., Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,J., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Louisedge,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
 Ogih,M., Okunou,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B.,  
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,  
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,  
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
 Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Woodson,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.  
 and Gibbs,R.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Direct Submission  
 Unpublished  
 2 (bases 1 to 150597)  
 Worley,K.C.  
 Direct Submission  
 Submitted (14-FEB-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Mar 4, 2001 this sequence version replaced gi:13162443.  
 ----- Genome Center  
 Center: Baylor College of Medicine

Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HCVU  
 Center clone name: RP11-112N23  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; M77789  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 123348 bases at least Q40  
 Consensus quality: 136820 bases at least Q30  
 Consensus quality: 142590 bases at least Q20  
 Estimated insert size: 124022; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 2.9x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 60 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 9261 9260: contig of 9260 bp in length  
 9361 9360: gap of unknown length  
 13390 13389: contig of 4029 bp in length  
 13489 13489: gap of unknown length  
 18686 18685: contig of 5196 bp in length  
 18786 18785: gap of unknown length  
 24700 24699: contig of 5914 bp in length  
 24700 24699: gap of unknown length  
 24800 24799: contig of 3882 bp in length  
 28682 28681: contig of 5305 bp in length  
 28782 28781: gap of unknown length  
 34087 34086: contig of 3712 bp in length  
 34187 34186: gap of unknown length  
 37899 37898: contig of 2923 bp in length  
 37999 37998: gap of unknown length  
 40922 40921: gap of unknown length  
 41022 41021: gap of unknown length  
 45839 45838: contig of 4817 bp in length  
 45939 45938: gap of unknown length  
 49512 49511: contig of 3573 bp in length  
 49612 49611: gap of unknown length  
 52526 52525: contig of 2915 bp in length  
 52627 52626: gap of unknown length  
 55606 55605: contig of 3879 bp in length  
 56606 56605: gap of unknown length  
 59514 59513: contig of 2908 bp in length  
 59614 59613: gap of unknown length  
 62615 62614: contig of 3001 bp in length  
 62715 62714: gap of unknown length  
 65936 65935: contig of 3221 bp in length  
 66036 66035: gap of unknown length  
 70197 70196: contig of 4161 bp in length  
 70297 70296: gap of unknown length  
 74089 74088: contig of 3792 bp in length  
 74189 74188: gap of unknown length  
 77996 77995: contig of 3807 bp in length  
 78096 78095: gap of unknown length  
 80751 80750: contig of 2655 bp in length  
 80851 80850: gap of unknown length  
 83712 83711: contig of 2861 bp in length  
 83812 83811: gap of unknown length  
 85968 85967: contig of 2157 bp in length  
 86069 86068: gap of unknown length  
 88107 88106: contig of 2038 bp in length  
 88207 88206: gap of unknown length  
 90490 90489: contig of 2283 bp in length  
 90589: gap of unknown length

90590 93053: contig of 2464 bp in length  
93054 93153: gap of unknown length  
93154 93224: contig of 2071 bp in length  
95325 95324: gap of unknown length  
95325 97203: contig of 1879 bp in length  
97204 97303: gap of unknown length  
97304 99435: contig of 2132 bp in length  
99436 99535: gap of unknown length  
99536 101546: contig of 2011 bp in length  
101547 101646: gap of unknown length  
101647 103981: contig of 2335 bp in length  
103982 104081: gap of unknown length  
104082 105715: contig of 1634 bp in length  
105716 105815: gap of unknown length  
105816 108300: contig of 2485 bp in length  
108301 108400: gap of unknown length  
108401 110338: contig of 1938 bp in length  
110339 110438: gap of unknown length  
110439 111900: contig of 1462 bp in length  
111901 112000: gap of unknown length  
112001 113693: contig of 1693 bp in length  
113694 113793: gap of unknown length  
113794 114955: contig of 1162 bp in length  
114956 115055: gap of unknown length  
115056 116317: contig of 1262 bp in length  
116318 116417: gap of unknown length  
116418 117742: contig of 1325 bp in length  
117743 117842: gap of unknown length  
117843 120255: contig of 2413 bp in length  
120256 120355: gap of unknown length  
120356 121589: contig of 1234 bp in length  
121590 121689: gap of unknown length  
121690 123233: contig of 1544 bp in length  
123234 123333: gap of unknown length  
123334 124567: contig of 1234 bp in length  
124568 124667: gap of unknown length  
124668 125989: contig of 1322 bp in length  
125990 126089: gap of unknown length  
126090 127484: contig of 1395 bp in length  
127485 127584: gap of unknown length  
127585 129477: contig of 1893 bp in length  
129478 129577: gap of unknown length  
129578 130833: contig of 1256 bp in length  
130834 130933: gap of unknown length  
130934 132130: contig of 1197 bp in length  
132131 132230: gap of unknown length  
132231 133460: contig of 1230 bp in length  
133461 133560: gap of unknown length  
133561 134771: contig of 1211 bp in length  
134772 134871: gap of unknown length  
134872 136312: contig of 1441 bp in length  
136313 136412: gap of unknown length  
136413 137636: contig of 1224 bp in length  
137637 137736: gap of unknown length  
137737 138779: contig of 1043 bp in length  
138780 138879: gap of unknown length  
138880 140120: contig of 1241 bp in length  
140121 141431: contig of 1211 bp in length  
140221 141531: gap of unknown length  
141432 141532: gap of unknown length  
141532 142750: contig of 1219 bp in length  
142751 142850: gap of unknown length  
142851 144113: contig of 1263 bp in length  
144114 144213: gap of unknown length  
144214 145697: contig of 1484 bp in length  
145698 145797: gap of unknown length

Query Match 11.0%; Score 325.6; DB 77; Length 150597;  
Best Local Similarity 73.5%; Pred. No. 4e-76;  
Matches 591; Conservative 2; Mismatches 161; Indels 50; Gaps 12;

QY 2137 caaagagcttgatgcaccctga-cagtgaggcttcgcatcctcatctccctca 2195  
|||||

Db 53420 CAAAGAGCTTGAGTTTATCACCCTTAGGGGTGCTCTCAGCACTTCATCATCT--CTTA 53363  
QY 2196 tcgtcaactactcatcaatgccaagctgatgtcctttcattgtctcccgagccag 2255  
Db 53362 TTGTACCTTACCGGATCCGTGCCAAA-----GCTCTAGAGCCAG 53321  
QY 2256 gagctgtgcatactgagsgagccscagctttcttgcagctcagcagttcactct 2315  
Db 53320 GAGCTGGGCTTAGTGAGGTGAGCCCGCAGCTTTGCCCTGCCAGCTTGCAATTCACCTCT 53261  
QY 2316 agagcatctgtccctcactgagacac-aaccactaattgtcactggaacctccctggcct 2374  
Db 53260 AGATCATTTGTTCCCTGGGACACAAACTACTAGTTTGTCACTGGAAGCTTCTGGGCT 53201  
QY 2375 gtcacagatgagtaacataaagagtggaactatccaaagagacagaggaataa 2434  
Db 53200 GTCTCAGAAATGGGATTAAAGAGAGTGGAAGTGTTCAGCAGAGGAGGAGAAATTA 53141  
QY 2435 ataattgcctccctcctcctccctccctccctccctccctccctccctccctccctcc 2494  
Db 53140 GTAAGTCCCTCC--TCTCTACCTTCTCTTCCAACTCCCTCCCTGCCATTTCTCTCT 53083  
QY 2495 ttctctactatgccaagatttgagattacaataagaagctctgtcctgtttaactc 2554  
Db 53084 ---TCTTCCCTCCCGGATTCACAGGCTTACAAATAGAGGCTTCTTCTCCGTTAATTC 53028  
QY 2555 cctagttaccacccctaattgtccctcagagaccttacttcttccctcgc-cctg 2613  
Db 53027 CATAGTACCACACCTAGTTTGCCCTTCTAGAAATCTCTACTCTCATATTGATCCG 52968  
QY 2614 tactctctctgtctctcaccaccacc-cgtacaccagcaacctcctgact--ggga 2669  
Db 52967 GTTCTGTACCTCTCCCGGATTCACAGGCTTACAAATAGAGGCTTCTTCTCCGTTAATTC 52908  
QY 2670 agacataaaggttaatgtcagggtaacactacatgagccct-----ga 2717  
Db 52907 AGGACAGGAAGTGTGAATGTACAGATCAAAATGTACAGATCCCTTCCCTGAGGAGA 52848  
QY 2718 ggcagaggcatctctgtgagctactgtctcctccactgtccttccactgtccttccagggc 2777  
Db 52847 GGATAGGGAGATCTCTGGGCTGAGGCTCTCTCTCTCCACTGCTTCTGAGGCTC 52788  
QY 2778 ctacagatgacacattaggtgtggcgtgt-gcggtgtgtatccaccctcagccca-- 2833  
Db 52787 CTCAGATGGCACACTAGGCTGGGCTTGTCTGTGGGTAGGATCCCACTCCAGTCCAGTG 52728  
QY 2834 cagtgctcagttgacttttataagctgtaatatctattttgttcttcttcc 2893  
Db 52727 GAATGCTTAGTGTAGCTCTTATTATGCTGTAAATGTTATTTTTAAGGGCGTACATATGA 52668  
QY 2894 ttattcttttgaatatatat 2917  
Db 52667 ATAGCAAGTTTGTGAATATAGAT 52644

RESULT 10  
AC074310 LOCUS AC074310 223431 bp DNA HTG 31-MAR-2001  
DEFINITION Mus musculus chromosome 1 clone RP23-137120, WORKING DRAFT  
ACCESSION AC074310 GI:13491228  
VERSION AC074310.18  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (Bases 1 to 223431)  
AUTHORS Rahhal, R., Yao, Z., Gu, W., and Roe, B. A.  
TITLE Mus musculus Chromosome 1 BAC Clone rp23-137120  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 223431)  
AUTHORS Rahhal, R., Yao, Z., Gu, W. and Roe, B. A.





\* be preserved

[illegible]

```
misc_feature /note="assembly_name:Contig16" 3367. .5850
/note="assembly_name:Contig17
clone_end:T7
vector_side:left"
5951. .8168
/note="assembly_name:Contig18"
8269. .10690
/note="assembly_name:Contig19"
10791. .13559
/note="assembly_name:Contig20"
13660. .16200
/note="assembly_name:Contig21"
16301. .19816
/note="assembly_name:Contig22"
19917. .23590
/note="assembly_name:Contig23"
23691. .27996
/note="assembly_name:Contig24"
28097. .31548
/note="assembly_name:Contig25"
31649. .35009
/note="assembly_name:Contig26"
35110. .39276
/note="assembly_name:Contig27"
39377. .43755
/note="assembly_name:Contig28"
43856. .49389
/note="assembly_name:Contig29"
49490. .53340
/note="assembly_name:Contig30"
53441. .58344
/note="assembly_name:Contig31"
58445. .64500
/note="assembly_name:Contig32"
64601. .70111
/note="assembly_name:Contig33"
70212. .77294
/note="assembly_name:Contig34"
77395. .83848
/note="assembly_name:Contig35"
83949. .90533
/note="assembly_name:Contig36"
90634. .96905
/note="assembly_name:Contig37"
97006. .103723
/note="assembly_name:Contig38"
103824. .112141
/note="assembly_name:Contig39"
112242. .120000
/note="assembly_name:Contig40"
120101. .129888
/note="assembly_name:Contig41"
129989. .141555
/note="assembly_name:Contig42"
141656. .153567
/note="assembly_name:Contig43"
153668. .165498
/note="assembly_name:Contig44"
165599. .176644
/note="assembly_name:Contig45"
176745. .190442
/note="assembly_name:Contig46"
33920 a 33920 c 34526 g 58493 t 3123 others
```

Db	104371	TGAGTTTGGC	ACTTGATGGTGGGCTTTGGCATCC	CATCTTCTCGTTGTC-ACTTATA	104313
QY	2206	ctgcatcaatg	cccaaaagctg	atgcttccttcattgctcccgaggagccaggaagctgtgc	2265
Db	104312	CTCCATCATATG	CACAAACTAATGTGTTTGTGCT-CCCTTGAGCCAGGAGGCTGTGTC		104254
QY	2266	atactgagsagg	aaccscaagcttctctgccaagctcagcaagttcaacttccttagagcatctg	2325	
Db	104253	TTACTAAAGAGG	ACCCTTAGCTTTCCTGTGCTGGCTCAGCAGTTCACTTCTAGATTATTT	104194	
QY	2326	tcccaactg	ggacacaaaccactaatgtgtcaactggaacctccctggcgctgtctcagaattg	2385	
Db	104193	TCCACACAGG	GACATTAACCACTAGTTGTGTCCACCGAAGCTCTGTTGGAACCTGCTCAGACTG	104134	
QY	2386	ggattacaataa	aaagatggaactataccaaaagagacagggagaataataaatgtgct	2445	
Db	104133	GGATTAAACAT	AAAGGAGTGGAACTGTACAAAAAGAG--AGAAAGAGATAAAGGAGGCTCCT	104076	
QY	2446	ccctccctccg	ctccc-----cttcccatcaacccttcccatlctcctctcct	2498	
Db	104075	CCCTTCCTCAG	CTCCCGGCTGTCTCCCTTCTGCTTTCACCTTTTTCCTTACCTT	104016	
QY	2499	ctactcatg	ccagatttgygattacaatagaagcttctgtcctgtltaactcccta	2558	
Db	104015	TCCAGATTCC	CAGATTCTGGGATTATGATAGAGACTTCTAGCTCCTGTTAACTCCTTA	103956	
QY	2559	gttaaccac	ccctaatttgccttc-----aggaaccttctaacttcttcctcctgacct	2612	
Db	103955	GTTACCCAC	CCCTCATTTGCCCTTCTCCAGGATCTCTTCTACTTTTTCATTTGTGCTT	103896	
QY	2613	gtacctctctg	ctctctcaaccaccacccctgttaaccagccacct	2657	
Db	103895	GTACCTCC	CTGTGCTTCCACCAACACCACTCTCTCTAGCTCCT	103851	

RESULT	12
AL138932/c	
LOCUS	
DEFINITION	
ACCESSION	AL138932
VERSION	170677 bp
KEYWORDS	Homo sapiens chromosome 1 clone RP1-244G5 map p11.1-13.3, ***
SOURCE	SEQUENCING IN PROGRESS ***, 31 unordered pieces.
ORGANISM	AL138932
	AL138932.10 GI:9230945
	HTG; HTGS_PHASE1; HTGS_DRAFT.
	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 170677)
AUTHORS	Pavitt,R.
TITLE	Direct Submission
JOURNAL	Submitted (13-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
	CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
	requests: clonerequest@sanger.ac.uk
COMMENT	On Jul 16, 2000 this sequence was submitted as part of a project to

```

web site: http://www.sanger.ac.uk
Contact: humphrey@sanger.ac.uk
----- Project Information -----
Center project name: dJ244G5
----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 39% of reads
Chemistry: Dye-terminator Big-amerham; 60% of reads
quality: 156012 bases at least Q40
Consensus quality: 160965 bases at least Q30
Consensus quality: 164137 bases at least Q20
Insert size: 167677; sum-of-contigs
Insert size: 131821; 35.2% error; agarose-fp
Quality coverage: 3.70x in Q20 bases; sum-of-contigs
Quality coverage: 4.70x in Q20 bases; agarose-fp

```

FEATURES  
SOURCE

NOTE: This is a 'working draft' sequence. It currently consists of 31 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1      4821: contig of 4821 bp in length
4822 4921: gap of 100 bp
4922 11645: contig of 6724 bp in length
11646 11745: gap of 100 bp
11746 13929: contig of 2184 bp in length
13930 14029: gap of 100 bp
14030 16875: contig of 2846 bp in length
16876 16975: gap of 100 bp
16976 20388: contig of 3413 bp in length
20389 20488: gap of 100 bp
20489 29256: contig of 8768 bp in length
29257 29356: gap of 100 bp
29357 30470: contig of 1114 bp in length
30471 30570: gap of 100 bp
30571 34082: contig of 3512 bp in length
34083 34182: gap of 100 bp
34183 49557: contig of 15375 bp in length
49558 49657: gap of 100 bp
49658 51777: contig of 2120 bp in length
51778 51877: gap of 100 bp
51878 53122: contig of 1245 bp in length
53123 53222: gap of 100 bp
53223 56791: contig of 3569 bp in length
56792 56891: gap of 100 bp
56892 61955: contig of 5064 bp in length
61956 62055: gap of 100 bp
62056 70347: contig of 8292 bp in length
70348 70447: gap of 100 bp
70448 77556: contig of 7109 bp in length
77557 77656: gap of 100 bp
77657 81110: contig of 3454 bp in length
81111 81210: gap of 100 bp
81211 82231: contig of 1021 bp in length
82232 82331: gap of 100 bp
82332 91668: contig of 9337 bp in length
91669 91768: gap of 100 bp
91769 112272: contig of 20504 bp in length
112273 112372: gap of 100 bp
112373 120545: contig of 8173 bp in length
120546 120645: gap of 100 bp
120646 140354: contig of 19709 bp in length
140355 140454: gap of 100 bp
140455 141779: contig of 1325 bp in length
141780 141879: gap of 100 bp
141880 143172: contig of 1293 bp in length
143173 143272: gap of 100 bp
143273 150964: contig of 7692 bp in length
150965 151064: gap of 100 bp
151065 152556: contig of 1492 bp in length
152557 152656: gap of 100 bp
152657 153707: contig of 1051 bp in length
153708 153807: gap of 100 bp
153808 154871: contig of 1064 bp in length
154872 154971: gap of 100 bp
154972 156380: contig of 1409 bp in length
156381 156480: gap of 100 bp
156481 160084: contig of 3604 bp in length
160085 160184: gap of 100 bp
160185 163849: contig of 3665 bp in length
163850 163949: gap of 100 bp
163950 170677: contig of 6728 bp in length.
Location/Qualifiers
1. 170677
/Organism="Homo sapiens"
```

[illegible]

```
misc_feature 143273. .150964
              /note="assembly_fragment:00171"
misc_feature 151065. .152556
              /note="assembly_fragment:00257"
misc_feature 152657. .153707
              /note="assembly_fragment:01027"
misc_feature 153808. .154871
              /note="assembly_fragment:01448"
misc_feature 154972. .156380
              /note="assembly_fragment:01559"
misc_feature 156481. .160084
              /note="assembly_fragment:01800"
misc_feature 160185. .163849
              /note="assembly_fragment:00119
              fragment_chain:6"
misc_feature 163950. .170677
              /note="assembly_fragment:01585
              fragment_chain:6
```

Query Match	7.2%;	Score 213.6;	DB 78;	Length 170677;
Best Local Similarity	98.2%;	Pred. No. 5.5e-46;		
Matches 216;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0;

[illegible]

Accession	Sequence	Position
QY	70 tgtcagtgtagcctagagagcgccctaacagacagagagcccaacggggcttcgcgtc	129
Db 166343	 tctcagtgtagcctagagagcgccctaacagacagagagcccaacggggcttcgcgtc	166284

QY	130	gcagagagc	caagatg	ctgcta	cgcga	ggggtg	ctctg	ggc	tcgac	cccg	gga	agtc	ggg	189
Db	166283	GCACAGAGC	CAGATG	GCTAC	GCGAG	GCGGTG	GCTCT	GCGG	CTGAC	CCCGG	AAGT	CGGG		166224

QY 190 tctccctcgccctctgtcttctcgcgtcctactagcaggt 229  
|||||  
Db 166223 tctccctcgcccttctgtcttctcgcgtcctactagcaggt 166184

RESULT	13
AL391417/c	
LOCUS	
DEFINITION	
AL391417	148540 bp
Homo sapiens chromosome 6 clone RP11-15J23,	DNA
PROGRESS ***, 45 unordered pieces.	HTG *** SEQUENCING IN

ACCESSION	AL391417
VERSION	AL391417.4
KEYWORDS	GI:9944084
SOURCE	HTG; HTGS_PHASE1.
	human.

ORGANISM	REFERENCE
Homo sapiens	1 (bases 1 to 148540)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo;	

**TITLE** Direct Submission  
**JOURNAL** Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
**COMMENT** requests: clonerquest@sanger.ac.uk  
On Aug 29, 2000 this sequence version replaced gl:9931024.

----- Genome Center  
Center: Sanger Centre

Center code: SC

Web site: [http://www.ck12.org](http://http://www.ck12.org)

Web site: <http://www.danbury.edu>

contact: [mlinquey@esangeli.ac.uk](mailto:mlinquey@esangeli.ac.uk)  
 protect: Information

----- Project Information -----

Center project name: DA15J23

----- Summary Statistics

Assembly program: XGAP4; version 4.3

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 1008 of reads

Consensus quality: 112851 bases at least Q40

Consensus quality: 127377 bases at least 030

Consensus quality:	126503 bases at least	020
Consensus quality:	126503 bases at least	020

Consensus quality: 130203 bases at 1000 g=



Insert size: 144140; sum-of-contigs  
Insert size: 165039; 32.7% error; agarose-fp  
Quality coverage: 1.71x in Q20 bases; sum-of-contigs Quality  
coverage: 2.30x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently  
consists of 45 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

```
1 3077: contig of 3077 bp in length
3078 3177: gap of 100 bp
3178 5415: contig of 2238 bp in length
5416 5515: gap of 100 bp
5516 8529: contig of 3014 bp in length
8530 8629: gap of 100 bp
8630 10965: contig of 2336 bp in length
10966 11065: gap of 100 bp
11066 15870: contig of 4805 bp in length
15871 15970: gap of 100 bp
15971 19571: contig of 3601 bp in length
19572 19671: gap of 100 bp
19672 21869: contig of 2198 bp in length
21870 21969: gap of 100 bp
21970 24727: contig of 2758 bp in length
24728 24827: gap of 100 bp
24828 30698: contig of 5871 bp in length
30699 30798: gap of 100 bp
30799 33751: contig of 2953 bp in length
33752 33851: gap of 100 bp
33852 36057: contig of 2206 bp in length
36058 36157: gap of 100 bp
36158 43283: contig of 7126 bp in length
43284 43383: gap of 100 bp
43384 45501: contig of 2118 bp in length
45502 45601: gap of 100 bp
45602 47671: contig of 2070 bp in length
47672 47771: gap of 100 bp
47772 50828: contig of 3057 bp in length
50829 50928: gap of 100 bp
50929 53130: contig of 2202 bp in length
53131 53230: gap of 100 bp
53231 55968: contig of 2738 bp in length
55969 56068: gap of 100 bp
56069 59340: contig of 3272 bp in length
59341 59440: gap of 100 bp
59441 61462: contig of 2022 bp in length
61463 61562: gap of 100 bp
61563 63584: contig of 2022 bp in length
63585 63684: gap of 100 bp
63685 65805: contig of 2121 bp in length
65806 65905: gap of 100 bp
65906 68895: contig of 2990 bp in length
68896 68995: gap of 100 bp
68996 76202: contig of 7207 bp in length
76203 76302: gap of 100 bp
76303 80091: contig of 3789 bp in length
80092 80191: gap of 100 bp
80192 86118: contig of 5927 bp in length
86119 86218: gap of 100 bp
86219 91581: contig of 5363 bp in length
91582 91681: gap of 100 bp
91682 94916: contig of 3235 bp in length
94917 95016: gap of 100 bp
95017 97304: contig of 2288 bp in length
97305 97404: gap of 100 bp
97405 101833: contig of 4429 bp in length
101834 101933: gap of 100 bp
101934 104031: contig of 2098 bp in length
104032 104131: gap of 100 bp
```

## FEATURES

## source

```
104132 106959: contig of 2828 bp in length
106960 107059: gap of 100 bp
107060 109230: contig of 2171 bp in length
109231 109330: gap of 100 bp
109331 111343: contig of 2013 bp in length
111344 111443: gap of 100 bp
111444 113704: contig of 2261 bp in length
113705 113804: gap of 100 bp
113805 116442: contig of 2638 bp in length
116443 116542: gap of 100 bp
116543 119288: contig of 2746 bp in length
119289 119388: gap of 100 bp
119389 121810: contig of 2422 bp in length
121811 121910: gap of 100 bp
121911 128837: contig of 6927 bp in length
128838 128937: gap of 100 bp
128938 132510: contig of 3573 bp in length
132511 132610: gap of 100 bp
132611 135417: contig of 2807 bp in length
135418 135517: gap of 100 bp
135518 137565: contig of 2048 bp in length
137566 137665: gap of 100 bp
137666 140029: contig of 2364 bp in length
140030 140129: gap of 100 bp
140130 143581: contig of 3452 bp in length
143582 143681: gap of 100 bp
143682 145770: contig of 2089 bp in length
145771 145870: gap of 100 bp
145871 148540: contig of 2670 bp in length.

Location/Qualifiers
1. 148540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-15J23"
/clone_1lb="RP11-11.1"
1. 3077
/note="assembly-fragment:00020"
3178. 5415
/note="assembly-fragment:00042"
5516. 8529
/note="assembly-fragment:00065"
8630. 10965
/note="assembly-fragment:00096"
11066. 15870
/note="assembly-fragment:00097"
15971. 19571
/note="assembly-fragment:00140"
19672. 21869
/note="assembly-fragment:00178"
21970. 24727
/note="assembly-fragment:00184"
24828. 30698
/note="assembly-fragment:00191"
30799. 33751
/note="assembly-fragment:00193"
33852. 36057
/note="assembly-fragment:00216"
36158. 43283
/note="assembly-fragment:00235"
43384. 45501
/note="assembly-fragment:00240"
45602. 47671
/note="assembly-fragment:00302"
47772. 50828
/note="assembly-fragment:00308"
50929. 53130
/note="assembly-fragment:00320"
53231. 55968
/note="assembly-fragment:00321"
56069. 59340
/note="assembly-fragment:00328"
59441. 61462
```

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

```
misc_feature /note="assembly_fragment:00353" 61563. .63584
misc_feature /note="assembly_fragment:00452" 63685. .65805
misc_feature /note="assembly_fragment:00461" 65906. .68895
misc_feature /note="assembly_fragment:00470" 68996. .76202
misc_feature /note="assembly_fragment:00519" 76303. .80091
misc_feature /note="assembly_fragment:00544" 80192. .86118
misc_feature /note="assembly_fragment:00550" 86219. .91581
misc_feature /note="assembly_fragment:00563" 91682. .94916
misc_feature /note="assembly_fragment:00610" 95017. .97304
misc_feature /note="assembly_fragment:00614" 97405. .10183
misc_feature /note="assembly_fragment:00616" 101934. .104031
misc_feature /note="assembly_fragment:00623"
```

Query Match	6.18;	Score 180.8;	DB 81;	Length 148540;
Best Local Similarity	72.98;	Pred. No. 3.7e-37;		
Matches 366;	Conservative	0;	Mismatches 102;	Indels 34;
				Gaps 9;

QY	2461	ccttcccatcaccccccttcccattcctcttcctctctctactcaatgcccagatttgga	2520
Db	26354	 CCTTTCTGTCCTTTCCTCTCCATATTCCCTCT - CCTACCCTGTAGATACTGGGA	26297
QY	2521	ttacaatatgaagctctctgtctccgtttaactccctagttaaccaacctaatltgccct	2580
Db	26296	 TTACAAGAANAAGCTCCCTGTGCTCGCTCAATGTCCACAGTCACTTAGTTAACCCT	26237
QY	2581	tcaaga--cccttctaacttltctcctcctgcctgtacctctctcgtcctacc---	2634
Db	26236	 TCATGATCCCTTCCACTACTCTTCATCCTGCTCTATACTCCCTTCTCCACCACCCGG	26177
QY	2635	--cccaacctgtaccaccagccacctctctgactggaag--gacataaaagyltaatyt	2690
Db	26176	 CACCACCCACGTATGCCAACACACCTTCCTTAACAGGAGAAAGACAGAAGGTTAATGC	26117
QY	2691	cagggtcaaac-----tacatlgaagccctygaga-----caggggcatctctgagctg	2739
Db	26116	 CAGAATCAAACATATGTACATGAACCTGTGAGGAGGAAGACAGGGGTGATTGCTGGGCTG	26057
QY	2740	agcctactgtctccttcccactgtccttcttccagggccctcagatgycacattaggtg	2799
Db	26056	 AGCCTGCTGTTCCCTGCCA---TCTTTCPCACAGGTCCTCAGATGCCATATCATCGTGG	26000
QY	2800	gcgtgctg-cgggtggtgatccccacctccagcccacagtgctcagtgfgactttlatla	2858
Db	25999	 GTGTGCTGCCGGGTGGGCATCCACACCTCCAGCCACAGTGTCAAGTTCACTTTTAGTG	25940
QY	2859	agctgtataatactatlttgtlittlgctlttcccttatcttlttgtaatatata	2918
Db	25939	 AGCTGTAAATATCTATTTTCTTGCTTCTTTCCATTTATTTT-----GTAAAAATG	25887
QY	2919	taatgagtltcattaaaaataga	2940
Db	25886	 TACTGAGTTTTATTAATAATATA	25865

RESULT	14
AL445926	
LOCUS	AL445926 167253 bp DNA PRI 20-JAN-2001
DEFINITION	Human DNA sequence from clone RP11-439I6 on chromosome 9, complete sequence.
ACCESSION	AL445926
VERSION	AL445926.5 GI:12329473
KEYWORDS	HTG.

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	Laird, G.	Direct Submission	Submitted (20-JAN-2001)
		1 (bases 1 to 167253)			Sanger Centre, Hinxton, Cambridgeshire,

**COMMENT**

On Jan 21, 2001 this sequence version replaced g1:12243916. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-43916 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-43916.

FEATURES	Location/Qualifiers
source	1. .167253

```

SOURCE
10180 a 32101 c 31806 g 53166 t
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-43916"
/clone_lib="RPCI-11.2"
BASE COUNT 50180 a 32101 c 31806 g 53166 t
ORIGIN

```

Query Match	6.1%;	Score 180.8;	DB 90;	Length 167253;
Best Local Similarity	72.9%;	Pred. No. 3.7e-37;		
Matches 366;	Conservative	0;	Mismatches 102;	Indels 34;
				Gaps 9.

[illegible]

Db	102632	AGCCTGCTGTTTCCCTTGCCA--TTCTTTCTCCAGGTCCTCAGATGCCATATCATCGTGG	102688
QY	2800	gcgtgctg-cgggtgggtatcccaacctccagccacagtgctcagttgtacttlltata	2858
Db	102689	GTGTGCTGCCGGGTGGGCATCCACCTCCAGCCACACAGTGTTCAGTTCACTTTTAGTG	102748
QY	2859	agctgtaataatactatcttctgttttttgccttttccttatacttttggtaataataata	2918
Db	102749	AGCTGTAATATCTATTTTTCTTCTTGCTTTTCTTTCCTATTATTTT-----GTAAAAATG	102801
QY	2919	taatgagtttcattcaataaataaga	2940
Db	102802	TACTGAGTTTTATTAAATATA	102823

RESULT 15  
AL590397/C

LOCUS	168537 bp	DNA	HTG	13-APR-2001
DEFINITION	Homo sapiens chromosome 9 clone RP11-376O21, *** SEQUENCING IN			
PROGRESS	***, 3 unordered pieces.			
ACCESSION	AF500307			

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AL590397	3	GI:13620466	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP	human.
				Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 168537)  
Plumb, B.  
Direct Submission  
Submitted (12-APR-2001) Sanger Centre Hinxton Cambridge

**COMMENT**

Submitted (12-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequests@sanger.ac.uk  
On Apr 12, 2001 this sequence version replaced gi:13568355.  
----- Genome Center

Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
-----  
Project Information  
Center project name: DA376021

```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 167948 bases at least Q40
Consensus quality: 168125 bases at least Q30
Consensus quality: 168249 bases at least Q20
Insert size: 168337; sum-of-contigs
Insert size: 160691; 19.0% error; agarose-fp
Quality coverage: 9.38x in Q20 bases; sum-of-contigs
Quality coverage: 9.83x in Q20 bases; agarose-fp

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

6100: contig of 6100 bp in length  
6101 6200: gap of 100 bp  
6201 61918: contig of 55718 bp in length  
61919 62018: gap of 100 bp  
62019 168537: contig of 106519 bp in length.  
Location: 300000-450000

## FEATURES

### source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-376021"
/clone_lib="RPCI-11.2"

```

```
misc_feature      1. .6100
                  /note="assembly_fragment:02399
                  fragment_chain:1
                  clone_end:T7
                  vector_side:left"
6201. .61918
                  /note="assembly_fragment:03300
                  fragment_chain:1"
62019. .168537
                  /note="assembly_fragment:00055
                  fragment_chain:1
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      53297 a 31342 c 32090 g 51608 t 200 others
ORIGIN
```

[illegible]

```
Search completed: June 27, 2001, 17:23:44
Job time: 10068 sec
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2001, 15:06:06 ; Search time 195.6 Seconds  
(without alignments)  
9466.676 Million cell updates/sec

Title: US-09-541-094-13

Perfect score: 2949  
Sequence: 1 tctgcagaattcgcgttcg.....attaataagattatccac 2949

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDS1/gcgdata/geneseq/NA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/NA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/NA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/NA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/NA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/NA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/NA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/NA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/NA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/NA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/NA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/NA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/NA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/NA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/NA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/NA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/NA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/NA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/NA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/NA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/NA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2948.2	100.0	2949	21	Human PAMP coding
2	1746.4	59.2	2250	21	Mouse PAMP coding
3	624.4	21.2	638	20	EST clone DN697.
4	424	14.4	424	21	Human secreted pro
5	357.8	12.1	473	21	EST sequence #2.
6	350.2	11.9	463	21	EST sequence #3.
7	337.2	11.4	481	21	EST sequence #4.
8	328	11.1	328	14	Human brain Expres
9	273.6	9.3	398	21	EST sequence #5.
10	208.2	7.1	422	21	EST sequence #1.
11	187	6.3	247	20	EST clone BB54. H

12	127.2	4.3	172	21	AAA37894
13	120.6	4.1	425	21	AAA37895
14	54.4	1.8	2942	21	AAA37887
15	49.8	1.7	53526	19	AA194101
16	49.8	1.7	53577	17	AA18551
17	49.8	1.7	53577	19	AA194108
18	45.4	1.5	4698	13	AAQ32370
19	45.4	1.5	4698	15	AAQ72475
20	45.4	1.5	4698	16	AA101176
21	45.4	1.5	4698	17	AA114692
22	45.4	1.5	4698	20	AA184111
23	45	1.5	936	22	AA158252
24	45	1.5	936	22	AA158254
25	45	1.5	936	22	AA158257
26	45	1.5	936	22	AA158259
27	45	1.5	936	22	AA158255
28	45	1.5	936	22	AA158252
29	43.6	1.5	2422	21	AA176112
30	43.4	1.5	936	22	AA158252
31	43.4	1.5	936	22	AA158254
32	43.4	1.5	936	22	AA158257
33	43.4	1.5	936	22	AA158259
34	43.4	1.5	936	22	AA158262
35	43.4	1.5	936	22	AA158255
36	43.4	1.5	1361	21	AA126399
37	42.8	1.5	244	22	AA158238
38	42.4	1.4	2070	21	AA176167
39	42	1.4	1756	21	AA198993
40	42	1.4	1761	20	AA120412
41	42	1.4	2198	21	AA123424
42	41.8	1.4	244	22	AA158238
43	41.6	1.4	3308	21	AA18270
44	41.6	1.4	5945	21	AA176030
45	41.4	1.4	289	14	AA053450

## ALIGNMENTS

RESULT 1	AAA37885	standard; DNA; 2949 BP.
ID	AAA37885	
XX	AAA37885;	
AC	12-FEB-2001	(first entry)
DT		
XX		
DE	Human PAMP coding sequence.	
XX		
KW	PAMP; human; presenilin associated membrane protein; immunogen;	
KW	neurodegenerative disease; Alzheimer's disease; Lewy body variant;	
KW	Parkinson's disease-dementia complex; neuropsychiatric disease;	
KW	schizophrenia; age-associated memory loss; developmental disorder;	
KW	neoplasm; diagnosis; ds.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	143..2272
FT		/*tag= a
FT		/product= PAMP
XX		
PN	WO200060069-A1.	
XX		
PD	12-OCT-2000.	
XX		
PF	03-APR-2000; 2000WO-CA00354.	
XX		
PR	01-APR-1999; 99US-0127452.	
PR	30-DEC-1999; 99US-0173826.	
XX		
PA	(UTOR ) UNIV TORONTO GOVERNING COUNCIL.	
XX		

EST sequence #6.  
EST sequence #7.  
D. melanogaster PA  
Human PKD1 gene.  
Human polycystic k  
Human PKD1 locus b  
PIA genomic DNA.  
Entire PIA genomic  
P815A antigen prec  
MAGE genomic DNA.  
PIA genomic sequen  
Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide D2  
Oligonucleotide D2  
Oligonucleotide D1  
Human ORF1667  
Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide D2  
Oligonucleotide D2  
Oligonucleotide D1  
Human secreted pro  
Human secreted pro  
Human ORF1722  
Human ORF1722  
Human pancreatic c  
Human secreted pro  
CDNA encoding huma  
Oligonucleotide D1  
Lung cancer associ  
Human ORF1585  
Sequence of the 3'





Db 1681 tgaaccacaacggtaccgcgtctctatgtgttcctcgtatlaaagccacaactcatg 1740  
QY 1741 gtccagctctatcctcagcagcagcactaaggctcacttggtgacggcctctcaaca 1800  
Db 1741 gtccagctctatcctcagcagcagcactaaggctcacttggtgacggcctctcaaca 1800  
QY 1801 ttacatcgctgtctccagcccccaacaaccactatgtgttacaglatgcttgcca 1860  
Db 1801 ttacatcgctgtctccagcccccaacaaccactatgtgttacaglatgcttgcca 1860  
QY 1861 ttgactggaacagtggttcaacctcaaccgagagcagtgccagatccaagttaagtc 1920  
Db 1861 ttgactggaacagtggttcaacctcaaccgagagcagtgccagatccaagttaagtc 1920  
QY 1921 aagtgaacaacagatctgtatgagtaactatggttcagagcccttgcaattcaatga 1980  
Db 1921 aagtgaacaacagatctgtatgagtaactatggttcagagcccttgcaattcaatga 1980  
QY 1981 gacggaccgactcccccggtgtgtgcttactgcaagattagccagggcctgtctcc 2040  
Db 1981 gacggaccgactcccccggtgtgtgcttactgcaagattagccagggcctgtctcc 2040  
QY 2041 tgccttgaactgagtcagtgagctctactgaatactctacatgagctgagagccgtg 2100  
Db 2041 tgccttgaactgagtcagtgagctctactgaatactctacatgagctgagagccgtg 2100  
QY 2101 gaaagatatccgtgcccgatattctctacatgcgcagcaaaagagcttgagtgatcac 2160  
Db 2101 gaaagatatccgtgcccgatattctctacatgcgcagcaaaagagcttgagtgatcac 2160  
QY 2161 gacagtggtgctcgagcactcactatcttctccctcatcgtcacactactgcatcaatg 2220  
Db 2161 gacagtggtgctcgagcactcactatcttctccctcatcgtcacactactgcatcaatg 2220  
QY 2221 agctgatgtccttctcactgtctcccgaggagccagagctgtgtcatactgagagacc 2280  
Db 2221 agctgatgtccttctcactgtctcccgaggagccagagctgtgtcatactgagagacc 2280  
QY 2281 scagcttctctgcccagctcagcaggttcaactcctcagagcatctgtcccaactgagaca 2340  
Db 2281 scagcttctctgcccagctcagcaggttcaactcctcagagcatctgtcccaactgagaca 2340  
QY 2341 accactaattgtcactggaacccctcctggtcctgtctcagattggatatacaataaag 2400  
Db 2341 accactaattgtcactggaacccctcctggtcctgtctcagattggatatacaataaag 2400  
QY 2401 agtgaactatccaaagagagcagggagaataaataattgctccctcctcctcc 2460  
Db 2401 agtgaactatccaaagagagcagggagaataaataattgctccctcctcctcc 2460  
QY 2461 ccttcccatcaccctcctccatctcctcctcctcctcctcctcctcctcctcctcctc 2520  
Db 2461 ccttcccatcaccctcctccatctcctcctcctcctcctcctcctcctcctcctcctc 2520  
QY 2521 ttacaataagaagctctgtcctgtttaactccctagttacccaaccctaattgacct 2580  
Db 2521 ttacaataagaagctctgtcctgtttaactccctagttacccaaccctaattgacct 2580  
QY 2581 tcagagcccttctacttctcctcctgacctgtacactctctcctcctcctcctcctcctc 2640  
Db 2581 tcagagcccttctacttctcctcctgacctgtacactctctcctcctcctcctcctcctc 2640  
QY 2641 cctgtaccagcaccctcctcctgactggtgaagacataaagtttaattgacaggtcaaa 2700  
Db 2641 cctgtaccagcaccctcctcctgactggtgaagacataaagtttaattgacaggtcaaa 2700  
QY 2701 ctacattgagccctgagagcagggcactctggtgagcctactgtctcctcctccac 2760  
Db 2701 ctacattgagccctgagagcagggcactctggtgagcctactgtctcctcctccac 2760  
QY 2761 tgccttctcctcagcctcagatgacacattaggggtggtgctgctgaggggtggtatcc 2820  
Db 2761 tgccttctcctcagcctcagatgacacattaggggtggtgctgctgaggggtggtatcc 2820

QY 2821 caactcagccacagtgctcagttgtacttttattaaagctgtaataatctatttgtt 2880  
Db 2821 caactcagccacagtgctcagttgtacttttattaaagctgtaataatctatttgtt 2880  
QY 2881 ttgtcttttctccttattcttttgttaataataataataatgagttcattcaataaga 2940  
Db 2881 ttgtcttttctccttattcttttgttaataataataataatgagttcattcaataaga 2940  
QY 2941 ttatccac 2949  
Db 2941 ttatccac 2949

RESULT 2  
AAA37886  
ID AAA37886 standard; DNA: 2250 BP.  
XX  
AC AAA37886;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Mouse PAMP coding sequence.  
XX  
KW PAMP; mouse; presenilin associated membrane protein; immunogen;  
KW neurodegenerative disease; Alzheimer's disease; Lewy body variant;  
KW parkinson's disease-dementia complex; neuropsychiatric disease;  
KW schizophrenia; age-associated memory loss; developmental disorder;  
KW neoplasm; diagnosis; ds.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 18..2144  
FT /tag= a  
FT /product= PAMP

WO200060069-A1.  
12-OCT-2000.  
XX  
PD 03-APR-2000; 2000MO-CA00354.  
PF  
XX  
PR 01-APR-1999; 99US-0127452.  
PR 30-DEC-1999; 99US-0173826.  
XX  
PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
PI St George-Hyslop PH, Fraser PE;  
XX  
DR WPI: 2000-665001/64.  
DR P-PSDB; AAY97550.  
XX  
PT Isolated presenilin associated membrane proteins and nucleic acids  
PT encoding them, useful for investigating and diagnosing Alzheimer's  
PT disease and other neurodegenerative diseases  
XX  
PS Claim 9; Page 70-71; 79pp; English.

This sequence encodes the mouse presenilin associated membrane protein to (PAMP) of the invention. PAMP polypeptides may be used as an immunogen to generate antibodies that recognise the PAMP polypeptide. The PAMP nucleotide and protein sequence may also be used for diagnosing individuals who are at risk or who have a variety of neurodegenerative diseases (e.g. Alzheimer's disease, Lewy body variant, Parkinson's disease-dementia complex), neuropsychiatric diseases (e.g. schizophrenia, age-associated memory loss), developmental disorders, and neoplasms. These may further be used to deduce the structural organisation and topology of PAMP, to identify proteins which interact with PAMP either in concert with presenilin 1 (PS1) and PS2, or independently, and to create cell-free systems, transfected cell lines, and animal models of neurodegenerative and other diseases.







XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1; SEQ ID 140; 71pp + CD-ROM; English.  
XX

CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
SO Sequence 424 BP; 98 A; 104 C; 121 G; 101 T; 0 other;

Query Match 14.4%; Score 424; DB 21; Length 424;  
Best Local Similarity 100.0%; Pred. No. 1.2e-107;  
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 agagaggaagatgctacgagcagggggtgctctggtgaccccggaagtcggggtc 191  
DB 1 agagaggaagatgctacgagcagggggtgctctggtgaccccggaagtcggggtc 60  
QY 192 tccttcgctctctgtcttctgctcctactagcaggtttgtgcaggggaaactcagttg 251  
DB 61 tccttcgctctctgtcttctgctcctactagcaggtttgtgcaggggaaactcagttg 120  
QY 252 agaggaagatataatatacccttaataaacaagctccctgtgttcgctgtcaacgcca 311  
DB 121 agaggaagatataatatacccttaataaacaagctccctgtgttcgctgtcaacgcca 180  
QY 312 ctcatcagattgctgctcagcttcaattagtgagacacaggggttatccacgtagtag 371  
DB 181 ctcatcagattgctgctcagcttcaattagtgagacacaggggttatccacgtagtag 240  
QY 372 agaaagaggaagacctaagtggtatagctatggtgcccccaacccctacatggttc 431  
DB 241 agaaagaggaagacctaagtggtatagctatggtgcccccaacccctacatggttc 300  
QY 432 tgctggagagcaagcatttaccaggagattatgagagaagctgaagaggagaacagcc 491  
DB 301 tgctggagagcaagcatttaccaggagattatgagagaagctgaagaggagaacagcc 360  
QY 492 gaattgctgtctgtcagtgctccttgacccaagccagctcctcaggtctctccta 551  
DB 361 gaattgctgtctgtcagtgctccttgacccaagccagctcctcaggtctctccta 420  
QY 552 gtgt 555  
DB 421 gtgt 424

RESULT 5  
AAA37890  
ID AAA37890 standard; DNA; 473 BP.

XX AAA37890;  
AC  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE EST sequence #2.  
XX  
KW PAMP; presenilin associated membrane protein; immunogen;

KW neurodegenerative disease; Alzheimer's disease; Lewy body variant;  
KW Parkinson's disease-dementia complex; neuropsychiatric disease;  
KW schizophrenia; age-associated memory loss; developmental disorder;  
KW neoplasm; diagnosis; EST; expressed sequence tag; ds.  
XX  
OS Unidentified.  
XX  
PN WO200060069-A1.  
XX  
PD 12-OCT-2000.  
XX  
PF 03-APR-2000; 2000WO-CA00354.  
XX  
PR 01-APR-1999; 99US-0127452.  
PR 30-DEC-1999; 99US-0173826.  
XX  
PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
XX  
PI St George-Hyslop PH, Fraser PE;  
XX  
DR WPI; 2000-665001/64.

PT Isolated presenilin associated membrane proteins and nucleic acids  
PT encoding them, useful for investigating and diagnosing Alzheimer's  
PT disease and other neurodegenerative diseases -  
XX  
PS Example 1; Page 60; 79pp; English.  
XX

CC This sequence represents an expressed sequence tag used to isolate the  
CC human presenilin associated membrane protein (PAMP) coding sequences  
CC of the invention. PAMP polypeptides may be used as an immunogen to  
CC generate antibodies that recognise the PAMP polypeptide. The PAMP  
CC nucleotide and protein sequence may also be used for diagnosing  
CC individuals who are at risk or who have a variety of neurodegenerative  
CC diseases (e.g. Alzheimer's disease, Lewy body variant, Parkinson's  
CC disease-dementia complex), neuropsychiatric disorders, and neoplasms.  
CC These may further be used to deduce the structural organisation and  
CC topology of PAMP, to identify proteins which interact with PAMP either  
CC in concert with presenilin 1 (PS1) and PS2, or independently, and to  
CC create cell-free systems, transfected cell lines, and animal models of  
CC neurodegenerative and other diseases.  
XX  
SO Sequence 473 BP; 129 A; 116 C; 117 G; 111 T; 0 other;

Query Match 12.1%; Score 357.8; DB 21; Length 473;  
Best Local Similarity 84.8%; Pred. No. 2.9e-89;  
Matches 401; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 249 tggagaggaagatataatcccttaataaacaagctccctgtgttcgctgtcaacg 308  
DB 1 tggagaggaagatataatcccttaataaacaagctccctgtgttcgctgtcaacg 60  
QY 309 ccactcatcagattgctgctcagcttcaattagtgagacacaggggttatccacgtag 368  
DB 61 ccactcatcagattgctgctcagcttcaattagtgagacacaggggttatccacgtag 120  
QY 369 tagagaaagagagacctaagtggtatgactgagtgcccaaccccttacatgg 428  
DB 121 tagagaaagagagacctaagtggtatgactgagtgcccaaccccttacatgg 180  
QY 429 ttctgtgagagcaagcatttaccaggagattatgagagaagctgaagaggagaacca 488  
DB 181 ttctgtgagagcaagcatttaccaggagattatgagagaagctgaagaggagaacca 240  
QY 489 gccgaattgctgtctcagtgctccttgacccaagccagctcctcaggtctctc 548  
DB 241 gtagaactgctgtctcagtgctccttgacccaagccagctcctcaggtctctc 300  
QY 549 ctagtgtacagtgcccaaatgatggtgtgttactccaattctctatggcagag 608  
DB 301 ctagtgtacagtgcccaaatgatggtgtgttactccaattctctatggcagag 360







XX WO200060069-A1.  
 PN 12-OCT-2000.  
 XX  
 PD  
 XX 03-APR-2000; 2000WO-CA00354.  
 PF  
 XX 01-APR-1999; 99US-0127452.  
 PR 30-DEC-1999; 99US-0173826.  
 XX  
 XX (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
 PA St George-Hyslop PH, Fraser PE;  
 PI WPI; 2000-665001/64.  
 DR

XX Isolated presenilin associated membrane proteins and nucleic acids  
 PT encoding them, useful for investigating and diagnosing Alzheimer's  
 PT disease and other neurodegenerative diseases -

XX Example 1; Page 61; 79pp; English.

XX This sequence represents an expressed sequence tag used to isolate the  
 CC human presenilin associated membrane protein (PAMP) coding sequences  
 CC of the invention. PAMP polypeptides may be used as an immunogen to  
 CC generate antibodies that recognise the PAMP polypeptide. The PAMP  
 CC nucleotide and protein sequence may also be used for diagnosing  
 CC individuals who are at risk or who have a variety of neurodegenerative  
 CC diseases (e.g. Alzheimer's disease, Lewy body variant, Parkinson's  
 CC disease-dementia complex), neuropsychiatric diseases (e.g. schizophrenia,  
 CC age-associated memory loss), developmental disorders, and neoplasms.  
 CC These may further be used to deduce the structural organisation and  
 CC topology of PAMP, to identify proteins which interact with PAMP either  
 CC in concert with presenilin 1 (PS1) and PS2, or independently, and to  
 CC create cell-free systems, transfected cell lines, and animal modes of  
 CC neurodegenerative and other diseases.

XX Sequence 398 BP; 107 A; 111 C; 99 G; 81 T; 0 other;

Query Match 9.3%; Score 273.6; DB 21; Length 398;  
 Best Local Similarity 83.5%; Pred. No. 6.2e-66;  
 Matches 334; Conservative 0; Mismatches 64; Indels 2; Gaps 2;

OY 1724 aaagccaactcattgttccagtctatcctcagcagcagcctaagtctctactgtgt 1783  
 DB 1 agagctaacaactcattgttccagtctatcctcagcagcagcctaagtctctactgtgt 59  
 OY 1784 gacgggcctctcaacattacatcgctgtctccagccccaacacacacttatgttga 1843  
 DB 60 gacagcctctcaacattacatcgctgtctccagccccaacacacacttatgttga 119  
 OY 1844 cagtatgccttggaatttgactgacagcagtggtcaacctcaccgagagcagtgccag 1903  
 DB 120 cagtatgccttggaatttgactgacagcagtggtcaacctcaccgagagcagtgccag 178  
 OY 1904 gatccaagtaaaagtcaccaagtgaaacaagagatctgtatgactcattggtccaggc 1963  
 DB 179 gatccaagtaaaagtcaccaagtgaaacaagagatctgtatgactcattggtccaggc 238  
 OY 1964 ccttgcatctaatgagacgagcagcagcagcagcagcagcagcagcagcagcagcagc 2023  
 DB 239 ccttgcatctaatgagacgagcagcagcagcagcagcagcagcagcagcagcagcagc 298  
 OY 2024 gccagggccttgctcctccttgcaactgagtcagtgagcagcagcagcagcagcagc 2083  
 DB 299 gccagggccttgctcctccttgcaactgagtcagtgagcagcagcagcagcagcagc 358  
 OY 2084 tggactgagagccgctggaagatatccgtgcccgatat 2123  
 DB 359 tggcgagagagccgctggaagatatccgtgcccgatat 398

RESULT 10

AAA37889

ID AAA37889 standard; DNA; 422 BP.

XX AAA37889;

XX 12-FEB-2001 (first entry)

XX EST sequence #1.

DE

XX PAMP; presenilin associated membrane protein; immunogen;  
 KW neurodegenerative disease; Alzheimer's disease; Lewy body variant;  
 KW Parkinson's disease-dementia complex; neuropsychiatric disease;  
 KW schizophrenia; age-associated memory loss; developmental disorder;  
 KW neoplasm; diagnosis; EST; expressed sequence tag; ds.

XX Unidentified.

XX WO200060069-A1.

XX 12-OCT-2000.

XX 03-APR-2000; 2000WO-CA00354.

XX 01-APR-1999; 99US-0127452.

XX 30-DEC-1999; 99US-0173826.

XX (UTOR ) UNIV TORONTO GOVERNING COUNCIL.

XX St George-Hyslop PH, Fraser PE;

XX WPI; 2000-665001/64.

XX Isolated presenilin associated membrane proteins and nucleic acids  
 PT encoding them, useful for investigating and diagnosing Alzheimer's  
 PT disease and other neurodegenerative diseases -

XX Example 1; Page 59; 79pp; English.

XX This sequence represents an expressed sequence tag used to isolate the  
 CC human presenilin associated membrane protein (PAMP) coding sequences  
 CC of the invention. PAMP polypeptides may be used as an immunogen to  
 CC generate antibodies that recognise the PAMP polypeptide. The PAMP  
 CC nucleotide and protein sequence may also be used for diagnosing  
 CC individuals who are at risk or who have a variety of neurodegenerative  
 CC diseases (e.g. Alzheimer's disease, Lewy body variant, Parkinson's  
 CC disease-dementia complex), neuropsychiatric diseases (e.g. schizophrenia,  
 CC age-associated memory loss), developmental disorders, and neoplasms.  
 CC These may further be used to deduce the structural organisation and  
 CC topology of PAMP, to identify proteins which interact with PAMP either  
 CC in concert with presenilin 1 (PS1) and PS2, or independently, and to  
 CC create cell-free systems, transfected cell lines, and animal modes of  
 CC neurodegenerative and other diseases.

XX Sequence 422 BP; 100 A; 92 C; 124 G; 106 T; 0 other;

Query Match 7.1%; Score 208.2; DB 21; Length 422;  
 Best Local Similarity 80.6%; Pred. No. 9e-48;  
 Matches 283; Conservative 0; Mismatches 58; Indels 10; Gaps 3;

OY 126 ctcagcagagagcagatgctacgagcaggggtgctctggtggtgacccgggaagtc 185  
 DB 1 ctcagcagagagcagatgctacgagcaggggtgctctggtggtgacccgggaagtc 60  
 OY 186 ggggtctccttcctctctctctctgctcctactagcaggttggtcaggggaact 245  
 DB 61 ggggtctccttcctctctctctctgctcctactagcaggttggtcaggggaact 117  
 OY 246 cagtggagaggaatataatcccttaataaacaagctccctgtgttcgctgtca 305  
 DB 118 cagtggagaggaatataatcccttaataaacaagctccctgtgttcgctgtca 177

OY 306 acgcacatcatcagatgtgctgccagttctcaattagtggagacacagggttatccacg 365  
|||||  
Db 178 acgcacatcatcagatgtgctgccagttctcaattagtggagacacagggttatccacg 237  
OY 366 tagtagaagaagaggagaccctacagtggttattgactgagcccccaacccccctaca 425  
|||||  
Db 238 tagtgagaagaagaaga-ctgaagtggtgtgttgacgagtgcccccaacccccctacatg 296  
OY 426 tggttctgtgagagcaagcatlcttaccagggtttaatgtgagaagctga 476  
|||||  
Db 297 gtctgtgtgaggggaag-----tcttcacagagatgtaatgtgagaagctga 341

## RESULT 11

AAV86828

ID AAV86828 standard; cDNA; 247 BP.

XX AAV86828;

DT 27-APR-1999 (first entry)

DE EST clone BB54.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN WO9845435-A2.

PD 15-OCT-1998.

PF 10-APR-1998; 98WO-US06954.

PR 10-APR-1997; 97US-0835913.

PA (GEM ) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;

DR WPI; 1999-070076/06.

PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries

PS Claim 1; Page 361; 633pp; English.

CC This sequence represents an expressed sequence tag (EST), and is a  
CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The EST sequences are also stated to be useful for gene  
CC therapy.

XX Sequence 247 BP; 62 A; 60 C; 53 G; 72 T; 0 other;

Query Match 6.3%; Score 187; DB 20; Length 247;  
Best Local Similarity 100.0%; Pred. No. 5e-42;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 578 ggtgttactccattctctatgggccagagtttctcactgcagagaataacagtggaat 637  
|||||  
Db 56 ggtgttactccattctctatgggccagagtttctcactgcagagaataacagtggaat 115  
OY 638 tcgctggcgaatggttggcttatgaagactttagttcccatcttcttcttgaagat 697  
|||||  
Db 116 tcgctggcgaatggttggcttatgaagactttagttcccatcttcttcttgaagat 175  
OY 698 gaaatgaaccacaagtcatacagcagtgctatcaagatcacacacctgagtcagaatgpc 757  
|||||  
Db 176 gaaatgaaccacaagtcatacagcagtgctatcaagatcacacacctgagtcagaatgpc 235  
OY 758 tcagcac 764  
|||||  
Db 236 tcagcac 242

## RESULT 12

AAA37894

ID AAA37894 standard; DNA; 172 BP.

XX AAA37894;

DT 12-FEB-2001 (first entry)

DE EST sequence #6.

KW PAMP; presenilin associated membrane protein; immunogen;  
neurodegenerative disease; Alzheimer's disease; Lewy body variant;  
Parkinson's disease-dementia complex; neuropsychiatric disease;  
schizophrenia; age-associated memory loss; developmental disorder;  
neoplasm; diagnosis; EST; expressed sequence tag; ds.

OS Unidentified.

PN WO200060069-A1.

PD 12-OCT-2000.

PF 03-APR-2000; 2000WO-CA00354.

PR 01-APR-1999; 99US-0127452.

PR 30-DEC-1999; 99US-0173826.

PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.

PI St George-Hyslop PH, Fraser PE;

DR WPI; 2000-665001/64.

PT Isolated presenilin associated membrane proteins and nucleic acids  
PT encoding them, useful for investigating and diagnosing Alzheimer's  
PT disease and other neurodegenerative diseases -

XX Example 1; Page 61-62; 79pp; English.

CC This sequence represents an expressed sequence tag used to isolate the  
CC human presenilin associated membrane protein (PAMP) coding sequences  
CC of the invention. PAMP polypeptides may be used as an immunogen to  
CC generate antibodies that recognise the PAMP polypeptide. The PAMP  
CC nucleotide and protein sequence may also be used for diagnosing  
CC individuals who are at risk or who have a variety of neurodegenerative  
CC diseases (e.g. Alzheimer's disease, Lewy body variant, Parkinson's  
CC disease-dementia complex), neuropsychiatric diseases (e.g. schizophrenia,  
CC age-associated memory loss), developmental disorders, and neoplasms.  
CC These may further be used to deduce the structural organisation and  
CC topology of PAMP, to identify proteins which interact with PAMP either  
CC in concert with presenilin 1 (PS1) and PS2, or independently, and to  
CC create cell-free systems, transfected cell lines, and animal models of  
CC neurodegenerative and other diseases.

SQ Sequence 172 BP; 37 A; 46 C; 50 G; 39 T; 0 other;





XX Isolated presenilin associated membrane proteins and nucleic acids  
PT encoding them, useful for investigating and diagnosing Alzheimer's  
PR disease and other neurodegenerative diseases -  
XX  
XX  
PS Claim 9; Page 73-75; 79pp; English.

CC This sequence encodes a presenilin associated membrane protein  
CC (PAMP) of the invention. PAMP polypeptides may be used as an immunogen to  
CC generate antibodies that recognise the PAMP polypeptide. The PAMP  
CC nucleotide and protein sequence may also be used for diagnosing  
CC individuals who are at risk or who have a variety of neurodegenerative  
CC diseases (e.g. Alzheimer's disease, Lewy body variant, Parkinson's  
CC disease-dementia complex), neuropsychiatric diseases (e.g. schizophrenia,  
CC age-associated memory loss), developmental disorders, and neoplasms.  
CC These may further be used to deduce the structural organisation and  
CC topology of PAMP, to identify proteins which interact with PAMP either  
CC in concert with presenilin 1 (PS1) and PS2, or independently, and to  
CC create cell-free systems, transfected cell lines, and animal models of  
CC neurodegenerative and other diseases.

XX  
SQ Sequence 2942 BP; 823 A; 729 C; 594 G; 796 T; 0 other;

Query Match 1.88; Score 54.4; DB 21; Length 2942;  
Best Local Similarity 44.7%; Pred. No. 0.00014;  
Matches 254; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 283 agtccctgtgttcgctcctcaagcaccatcagattgtgtcgcagcttcaatag 342  
DB 193 agctagctgttccgagcgtgaatgcccacatcagacagctgttcccaactc 252  
QY 343 tggagacacaggggtatccacgtagtagaagaagagagacacagctggtatgac 402  
DB 253 cgggtccgtgtggtactacataataacgtgcagcgcagctgtgaattcttctag 312  
QY 403 tgatgcccccaacccctacatgttctgtcgtgagagcaagcatttaccaggtatt 462  
DB 313 cagccaccatctccacttacgcccccatgataccacactcctgttccacagtaaca 372  
QY 463 aatgagaagctgaagagaggaacacagcgaattgtgtcttgcagtgcttgcacaa 522  
DB 373 cctgatgcgtcaaggaagccgagcaacgaacatttctgtgtgtgtgatcaaccg 432  
QY 523 gccagtcctcgtcctcagcgttctctcctagtgtacagtgcccccaatgtggtgt 582  
DB 433 cagcaaccagatgaagcagttctcgcagcaactcaactgcccccaatcagtaagcg 492  
QY 583 ttactccaattcctatggtcagcaggtttgtcactgcagagaaatacagtggaattc 642  
DB 493 gaacagca--ccagtgaagacctgcgaagcagcaatccagcaaaaaaactggaatcc 549  
QY 643 gggcaatgtgttggtatgaagactttagttcccatcttcttcttgaagatga 702  
DB 550 gggcactggaattctgcagcaggaacttcccttccatctatcatacagcagatttga 609  
QY 703 tgaacccaagtcatacagcagtgctatacacaacctgagtcagaatggtcagc 762  
DB 610 tcaggtcaccagaagtagaagagtgcttccaggaacttaacaacataactagagagc 669  
QY 763 accaaccctcccaactatgtgcagtcagctcttccacacatgcatgtgtcatcagc 822  
DB 670 cggcgtgctgtagctgtgtgcgcgtcagagtcagtaagtccttattgtccgcgtgt 729  
QY 823 tggcaactgcatgctgcgcagctcacc 850  
DB 730 cgaggtctgtatgctgcgcacacactc 757

RESULT 15  
AAT94101  
ID AAT94101 standard; DNA; 53526 BP.  
XX

AC AAT94101;  
XX  
XX  
DT 01-JUN-1998 (first entry)  
XX  
XX  
DE Human PKD1 gene.  
XX  
KW Human; polycystic kidney disease 1; PKD1; treatment;  
KW autosomal dominant polycystic kidney disease; APKD; ss.  
XX  
OS Homo sapiens.  
XX  
PN W09744457-A1.  
XX  
PD 27-NOV-1997.  
XX  
PF 22-MAY-1997; 97WO-US08799.  
XX  
PR 03-JUN-1996; 96US-0658136.  
PR 24-MAY-1996; 96US-0655360.  
XX  
XX

(GENZ ) GENZYME CORP.

PI Burn T, Connors T, Dackowski W, Germino G, Klinger K;  
PI Qian F;  
XX  
XX  
DR WPI; 1998-018511/02.  
XX  
XX

Human polycystic kidney disease gene, PKD1 - useful to treat and  
diagnose human autosomal or adult onset polycystic kidney disease;  
Claim 2; Pages 90-118; 257pp; English.

XX The present sequence is the human polycystic kidney disease 1  
CC (PKD1) gene. The PKD1 gene or polypeptide may be used to treat  
CC autosomal dominant polycystic kidney disease (APKD), and identify  
CC carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.  
CC Antibodies (Ab) that distinguish between normal and mutant PKD1  
CC sequences can also be used in diagnostic tests. Anti-PKD1 Ab may  
CC also be used to perform subcellular and histochemical localisation  
CC studies, and to block the function of PKD1. Ab are also useful in  
CC rational drug design studies to identify and test inhibitors of  
CC PKD1. Sense and antisense sequences derived from the PKD1 gene may  
CC used for detection and therapy.

XX  
SQ Sequence 53526 BP; 8486 A; 17665 C; 15768 G; 11607 T; 0 other;

Query Match 1.7%; Score 49.8; DB 19; Length 53526;  
Best Local Similarity 56.2%; Pred. No. 0.015;  
Matches 114; Conservative 0; Mismatches 87; Indels 2; Gaps 1;

QY 2441 tgcctccctcctcgtcgtcccttcccatcacccttcccatcttctctct 2500  
DB 35695 tccctcccttctctcttcttcttctccttcccttctcctcgttctccttcc 35754  
QY 2501 actcatgccaagatttgggattacaagaagcttctgtcctcgtttaaactcag 2560  
DB 35755 tctc--cccttctcctcctcctccttctcctcctcctccttcttcttctt 35812  
QY 2561 taccaaccataatttgccttccagcacccttctacttttctcctcctcgtgacct 2620  
DB 35813 ttccctccacttcccttcttcttcccttctccttctccttctccttctcct 35872  
QY 2621 ctctgctcctcaccaccacct 2643  
DB 35873 ctcttcttctccttctcctt 35895

Search completed: June 27, 2001, 16:53:52  
Job time: 6466 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2001, 14:51:51 ; Search time 86.97 Seconds  
(without alignments)  
6281.853 Million cell updates/sec

Title: US-09-541-094-13

Perfect score: 2949

Sequence: 1 tctgcgaattcgcgttcg.....atlaaatagattatccac 2949

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	79.2	2.7	7218	1 US-08-232-463-14	Sequence 14, Appl
2	49.8	1.7	53526	3 US-08-1658-136-2	Sequence 2, Appl
3	49.8	1.7	53577	3 US-08-658-136-1	Sequence 1, Appl
4	45.4	1.5	4698	1 US-07-807-043B-5	Sequence 5, Appl
5	45.4	1.5	4698	1 US-08-299-849B-5	Sequence 5, Appl
6	45.4	1.5	4698	2 US-08-142-368A-5	Sequence 5, Appl
7	45.4	1.5	4698	3 US-08-967-727-5	Sequence 5, Appl
8	45.4	1.5	4698	4 US-08-037-230D-5	Sequence 5, Appl
9	41.4	1.4	289	1 US-08-341-568-3	Sequence 3, Appl
10	41.4	1.4	289	2 US-08-911-020-3	Sequence 3, Appl
11	41.2	1.4	5852	1 US-07-867-106-2	Sequence 2, Appl
12	40.6	1.4	2409	4 US-09-293-322C-8	Sequence 8, Appl
13	40.2	1.4	19124	2 US-08-487-826B-13	Sequence 13, Appl
14	39.8	1.3	1117	4 US-09-247-373B-33	Sequence 33, Appl
15	39.6	1.3	1736	4 US-09-182-816-22	Sequence 22, Appl
16	39.6	1.3	1736	4 US-09-182-816-24	Sequence 24, Appl
17	39.6	1.3	1736	4 US-09-471-528-22	Sequence 22, Appl
18	39.6	1.3	1736	4 US-09-471-528-24	Sequence 24, Appl
19	39.4	1.3	11517	1 US-07-920-281C-1	Sequence 1, Appl
20	39.4	1.3	11517	4 US-08-466-277-1	Sequence 1, Appl
21	39.2	1.3	1582	3 US-08-545-196B-10	Sequence 10, Appl
22	39.2	1.3	1582	3 US-08-545-196B-12	Sequence 12, Appl
23	39	1.3	3848	4 US-09-112-096-28	Sequence 28, Appl
24	39	1.3	5668	4 US-09-112-096-14	Sequence 14, Appl
25	38.8	1.3	1641	1 US-08-300-903A-8	Sequence 8, Appl
26	38.4	1.3	960	3 US-09-248-335-57	Sequence 57, Appl
27	38.2	1.3	2852	3 US-09-027-137-2	Sequence 2, Appl

C 28	38	1.3	688	6	5498694-3	Patent No. 5498694
C 29	38	1.3	1493	1	US-08-340-820-24	Sequence 24, Appl
C 30	38	1.3	1493	1	US-08-593-535-24	Sequence 24, Appl
C 31	37.6	1.3	140	1	US-08-628-417-5	Sequence 5, Appl
C 32	37.6	1.3	240	1	US-08-628-417-6	Sequence 6, Appl
C 33	37.6	1.3	2294	4	US-08-964-700A-1	Sequence 1, Appl
C 34	37.6	1.3	7286	4	US-09-331-581-3	Sequence 3, Appl
C 35	37.6	1.3	7938	4	US-08-332-766A-16	Sequence 16, Appl
C 36	37.4	1.3	445	2	US-08-332-766A-16	Sequence 16, Appl
C 37	37.4	1.3	746	4	US-09-013-810-1	Sequence 1, Appl
C 38	37.4	1.3	2230	4	US-08-378-313-24	Sequence 24, Appl
C 39	37.4	1.3	3238	5	PCT-US94-10080-5	Sequence 5, Appl
C 40	37.4	1.3	5173	1	US-08-242-677-1	Sequence 1, Appl
C 41	37.4	1.3	5852	1	US-07-867-106-2	Sequence 2, Appl
C 42	37.2	1.3	966	2	US-08-766-738-2	Sequence 2, Appl
C 43	37	1.3	1474	4	US-08-821-994-64	Sequence 64, Appl
C 44	36.8	1.2	1441	4	US-08-821-994-63	Sequence 63, Appl
C 45	36.4	1.2	271	2	US-08-731-272A-29	Sequence 29, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZgpt-Fls  
US-08-232-463-14

```

; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 53577 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
US-08-658-136-1

```





ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5925729man D.

QY	2449	tctccgcgctccccctttcccatacaccoccttccccaatttcctctctctactcatgc	2508
Db	2559	TTCCTCCACCCTGCTTCTTTAACCTGCCTCCTCCCAATTGCCCTTACCTTTATGCCCATTC	2618
QY	2509	cagaatttgagatcacaaatagaagcttctgtctctgtttaactccoctagttaaccac	2568
Db	2619	CATGTCCCCTCTCAATTCCTCTGCCAATTGTGCTCCCTCACACTCTCCATTTCCCTCTT	2678
QY	2569	ctaattggcccttcaggaaccttctacttttcccttcctgcgccctgtactctctctgc	2628
Db	2679	CTCCCTTAGCCCTCTTCCTTCTCTCTTGATCTCCCTTCCTTTGGCTTCTCCCTCTC	2738
QY	2629	ctaccccccaacctgtatacccaagcaacctctctga	2663
Db	2739	CTTCCCCCTTCCCTATGCCCTTACTCTACTTGA	2773

:  
:  
: APPLICANT: Gaugler, B attrice; Van den Eynde, Beno t;  
:  
: TITLE OF INVENTION: van der Bruggen, Pierre; Boon-Falleur, Thierry  
:  
: TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For  
:  
: NUMBER OF SEQUENCES: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof  
:  
: CORRESPONDENCE ADDRESS: 30  
:  
:

```

? ZIP: 10022
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
?
? Commented: 170

```

```

;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;      CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/967,727

FILED DATE  
CLASSIFICATION

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037 330

```

FITTING DATE

FILING DATE: 28-MARCH-1993  
APPLICATION NUMBER: PCT/US92/04354  
FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/807,043  
FILING DATE: 12-DECEMBER-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/764,365  
FILING DATE: 23-SEPTEMBER-1991

```

; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6025474man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
;
US-08-667-727-5

```

Query Match	1.58;	Score 45.4;	DB 3;	Length 4698;
Best Local Similarity	50.78;	Pred. No. 0.011;		
Matches 109; Conservative	0;	Mismatches 106;	Indels 0;	Gaps 0;

QY	2449	ttctctcgcgtccccctttcccatcaccccccttcccccatlctctctctctctctactcatgc	2508
	2559	TTCTTCCACACCTGCTTCTTTACCCTGCGCTCTCCCAATGCCCTCTTACCTTTATGCCCAFTC	2618
QY	2509	cagattttggygatlacaataagaagcttctgtctcgtgttaactccctagttaaccacc	2568
Db	2619	CATGTCCCTCTCAATTCCTGTGCCAATTGTGTCCTCCACACATCTTCCATTTCCCTCTTT	2678
QY	2569	ctaatttgcccttcagagacccttctacttttctctctctgcccctgtaacctctctgctc	2628
Db	2679	CTCCCTTAGCCCTCTCTTCTTCCTCTCTCTGTGTATCCCTTCCCTTTGCTTCTCCCTCTCTC	2738
QY	2629	ctcaccccccaacccctgtatacccaagccacacctctga	2663
Db	2739	CTTTCCTCTTCCCTTAGGCCCTCTACTCTACTTGA	2773

RESULT 8  
US-08-037-230D-5  
; Sequence 5, Application US/08037230D  
; Patent No. 6235525

APPLICANT: Gauglier, B atrice; Van den Eynde, Beno t;  
 APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry  
 TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For  
 TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof  
 NUMBER OF SEQUENCES: 30  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Felfe & Lynch  
 STREET: 805 Third Avenue  
 CITY: New York City  
 STATE: New York

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/037, 230D  
FILING DATE: 26-MARCH-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/04354  
FILING DATE: 22-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/807,043

```

1      FILING DATE: 12-DECEMBER-1991
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: 07/764,364
4      FILING DATE: 23-SEPTEMBER-1991
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: 07/728,838
7      FILING DATE: 9-JULY-1991
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: 07/705,702
10     FILING DATE: 23-MAY-1991
11     ATTORNEY/AGENT INFORMATION:
12     NAME: Hanson, No. 6235525man D.
13     REGISTRATION NUMBER: 30,946
14     REFERENCE/DOCKET NUMBER: LUD 5353
15     TELECOMMUNICATION INFORMATION:
16     TELEPHONE: (212) 688-9200
17     TELEFAX: (212) 838-3884
18     INFORMATION FOR SEQ ID NO: 5:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 4698 base pairs
21     TYPE: nucleic acid
22     STRANDEDNESS: single
23     TOPOLOGY: linear
24     MOLECULE TYPE: genomic DNA
25
26     US-08-037-230D-5

```

Query Match	1.5%;	Score 45.4;	DB 4;	Length 4698;	Gaps
Best Local Similarity	50.7%;	Pred. No. 0.011;			0;
Matches 109;	Conservative	0;	Mismatches 106;	Indels 0;	

QY	2449	ttcttcgcgtcccccctttcccaataacccccttcccccacttctctctctctctactcatgc	2508
Db	2559	TTCTTCCACCTGCTTCTTTAACCTGCCTCTCCCATTTGCCCTTTACCTTTATGCCCATTC	2618
QY	2509	cagattttgggattacaataagaagcttcctgtcctcgtttaaactccctagttaccacc	2568
Db	2619	CATGTCCCTCTCAATTTCCCTGTCCCATTTGTGCTCCCTCACAATCTTCCATTTCCCTCTTT	2678
QY	2569	ctaatttgcccttcagagaccctctacttttctcctcctgcccctgtacactctctgctc	2628
Db	2679	CTCCCTTAGCCTCTTCTTCCCTCTCTGTATCTCCCTTTGCCCTTGTCTTCCCTCTTC	2738
QY	2629	ctaccccccaaccctgttaccccaagccacttctctga	2663
Db	2739	CTTTCCCTTCTCCCTATATGCCCTTACTCTACTTGA	2773

RESULT 9  
US-08-341-568-3/c  
; Sequence 3, Application US/08341568  
; Patent No. 5661021

APPLICANT: Buchert, Johanna  
 APPLICANT: Silka-aho, Matti  
 APPLICANT: Viikari, Liisa  
 APPLICANT: Penttila, Merja  
 APPLICANT: Saloheimo, Anu  
 APPLICANT: Marjatta, Rannu  
 TITLE OF INVENTION: Mannanase enzymes, genes coding for them,  
 TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching  
 TITLE OF INVENTION: pulps  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Birch, Stewart, Kolasch and Birch  
 STREET: PO Box 747  
 CITY: Falls Church  
 STATE: VA  
 COUNTRY: US  
 ZIP: 22040-0747  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible



APPLICATION NUMBER: US 08/341,568

TELEFAX: 213-268-3439  
INFORMATION FOR SEQ ID NO: 2

```

; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbie Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Query Match          1.4%; Score 40.2; DB 2; Length 19124;
Best Local Similarity 62.4%; Pred. No. 0.7;
Matches 63; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY   2844 ttgtacttttataagctgtaatatcatttggtttgccttttccttatctt 2903
      ||||| ||||||| | ||||| | ||||| ||||| ||||| ||||| |||||
DB   18385 TTATATTTTTATAATTATTAATTAATAAATTTTTTTTTTTTTTTTGTG 18326

QY   2904 ttgtaaatatataataatgatgttcattcaaaatagattat 2944
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   18325 ATATGTAGAACTACTTTATGTTTGGTTTGATGAATAAAGAAT 18285

RESULT 14
US-09-247-373B-33/c
; Sequence 33, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247,373B
; CURRENT FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924,747
; PRIOR FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 56

```

Search completed: June 27, 2001, 16:52:04  
Job time: 7213 sec

```

; SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
; LENGTH: 1117
; TYPE: DNA
; ORGANISM: SOYBEAN
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1101)
; OTHER INFORMATION: M=A OR C
; NAME/KEY: unsure
; LOCATION: (1104)
; OTHER INFORMATION: M=A OR C
; NAME/KEY: unsure
; LOCATION: (1116)
; OTHER INFORMATION: N=G OR A OR T OR C
US-09-247-373B-33

```

Query Match	1.38;	Score 39.8;	DB 4;	Length 1117;
Best Local Similarity	62.68;	Pred. No. 0.19;		
Matches 62; Conservative	0;	Mismatches 37;	Indels 0;	Gaps 0;

[illegible]

QY	2904	tcgtaaatatataataatgagttcattcaataaatagatt	2942
Dp	1012	TTTTTTTTTTTAAAGAAATTGAATTATGATTATTT	974

## RESULT 15

US-09-182-816-22/c  
; Sequence 22, Application US/09182816  
; Patent No. 6143542

```

GENERAL INFORMATION:
APPLICANT: wisniewski, Nancy
APPLICANT: Silver, Gary M.
APPLICANT: Lo, Katherine C.
APPLICANT: Brandt, Kevin S.
TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES
TITLE OF INVENTION: PROTEINS AND USES THEREOF
FILE REFERENCE: FC-3-C1
CURRENT APPLICATION NUMBER: US/09/182,816
CURRENT FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: 08/989,510
EARLIER FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 22
LENGTH: 1736
TYPE: DNA
ORGANISM: Ctenocephalides felis
FEATURE:
NAME/KEY: CDS
LOCATION: (159)..(1553)
US-09-182-816-22

```

Query Match	1.3%;	Score 39.6;	DB 4;	Length 1736;
Best Local Similarity	61.8%;	Pred. No. 0.27;		
Matches 63;	Conservative 0;	Mismatches 39;	Indels 0;	Gaps 0.

QY	2844	ttgtactcttctaagcgtgaatacatatttggtttgctctttccottatcttc                                      TTT	2903
Db	1727	TTT 	1668

QY	2904	ttgtaaatatatataatgatgttcattaaatagattatc	2945
Db	1667	TTTTTTTTTATATTTGACACAGTTTGAATAAATCCATATC	1626



Thu Jun 28 10:23:27 2001

us-09-541-094-13.rni



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2001, 14:35:06 ; Search time 2158.27 Seconds  
(without alignments)  
12916.116 Million cell updates/sec

Title: US-09-541-094-13  
Perfect score: 2949  
Sequence: 1 tctgcagaattcgcgttcg.....attaataatagattatccac 2949

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est110:\*  
191: gb\_est111:\*  
192: gb\_est112:\*  
193: gb\_est113:\*  
194: gb\_est114:\*  
195: gb\_est115:\*  
196: gb\_est116:\*  
197: gb\_est117:\*  
198: gb\_est118:\*  
199: gb\_est119:\*  
200: gb\_est120:\*  
201: gb\_est121:\*  
202: gb\_est122:\*  
203: gb\_est123:\*  
204: gb\_est124:\*  
205: gb\_est125:\*  
206: gb\_est126:\*  
207: gb\_est127:\*  
208: gb\_est128:\*  
209: gb\_est129:\*  
210: gb\_est130:\*  
211: gb\_est131:\*  
212: gb\_est132:\*  
213: gb\_est133:\*  
214: gb\_est134:\*  
215: gb\_est135:\*  
216: gb\_est136:\*  
217: gb\_est137:\*  
218: gb\_est138:\*  
219: gb\_est139:\*  
220: gb\_est140:\*  
221: gb\_est141:\*  
222: gb\_est142:\*  
223: gb\_est143:\*  
224: gb\_est144:\*  
225: gb\_est145:\*  
226: gb\_est146:\*  
227: gb\_est147:\*  
228: gb\_est148:\*  
229: gb\_est149:\*  
230: gb\_est150:\*  
231: gb\_est151:\*  
232: gb\_est152:\*  
233: gb\_est153:\*  
234: gb\_est154:\*  
235: gb\_est155:\*  
236: gb\_est156:\*  
237: gb\_est157:\*  
238: gb\_est158:\*  
239: gb\_est159:\*  
240: gb\_est160:\*  
241: gb\_est161:\*  
242: gb\_est162:\*  
243: gb\_est163:\*  
244: gb\_est164:\*  
245: gb\_est165:\*  
246: gb\_est166:\*  
247: gb\_est167:\*  
248: gb\_est168:\*  
249: gb\_est169:\*  
250: gb\_est170:\*  
251: gb\_est171:\*  
252: gb\_est172:\*  
253: gb\_est173:\*  
254: gb\_est174:\*  
255: gb\_est175:\*  
256: gb\_est176:\*  
257: gb\_est177:\*  
258: gb\_est178:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	827	28.0	840	108	AUI33372	AUI33372 AUI33372
2	797	27.0	853	154	BG489641	BG489641 602518064
3	768.2	26.0	838	108	AUI32380	AUI32380 AUI32380
4	753.8	25.6	784	19	AI377767	AI377767 te62b05.x
5	691.4	23.4	748	165	BE262466	BE262466 601154772
6	677.8	23.0	914	147	BF339787	BF339787 602034987
7	666.2	22.6	696	155	BG568503	BG568503 602587478
8	661.4	22.4	910	147	BF340237	BF340237 602036827
9	652.8	22.1	807	167	BE408912	BE408912 601303794
10	634.8	21.5	950	172	BF980176	BF980176 602288192
11	633.2	21.5	656	169	BF793556	BF793556 602255077
12	628.6	21.3	693	150	BF570121	BF570121 602185971
13	627	21.3	934	150	BF530600	BF530600 602071875
14	623	21.1	813	150	BF569485	BF569485 602185971
15	606.4	20.6	820	140	BE788789	BE788789 601475883
16	597.2	20.3	622	143	BF035070	BF035070 601456171
17	596.8	20.2	920	172	BF979552	BF979552 602288192
18	593	20.1	779	107	AUI25699	AUI25699 AUI25699
19	591.6	20.1	808	152	BG325507	BG325507 602424205
20	587.2	19.9	604	147	BF375675	BF375675 RC5-ST030
21	583.8	19.8	859	168	BF680467	BF680467 602154273
22	580.2	19.7	678	32	AV648359	AV648359 AV648359
23	580.2	19.7	639	115	AW372837	AW372837 RC1-BT037
24	580.2	19.7	639	115	BE907949	BE907949 601497312
25	580.2	19.7	916	141	BE307949	BE307949 601497312
26	576.4	19.5	618	115	AW386734	AW386734 RC0-PT002
27	574.4	19.5	602	115	AW368457	AW368457 CM4-HT019
28	573.2	19.4	913	141	BE909665	BE909665 601501673
29	572.4	19.4	598	143	BF035344	BF035344 601458551
30	570.4	19.3	584	141	BE890486	BE890486 601431556
31	565.8	19.2	670	167	BE388123	BE388123 601284450
32	561.8	19.1	610	152	BG325127	BG325127 602423676
33	560.2	19.0	655	32	AV699416	AV699416 AV699416
34	558.2	18.9	627	107	AUI28443	AUI28443 AUI28443
35	557.8	18.9	596	105	AL039414	AL039414 DKFZp434L
36	555	18.8	600	115	AW369337	AW369337 QV1-BN000
37	548	18.6	604	107	AUI27220	AUI27220 AUI27220
38	544.8	18.5	602	141	BE878476	BE878476 601487888
39	531.8	18.0	546	120	AW805004	AW805004 QV4-UM009
40	527.6	17.9	554	115	AW363805	AW363805 RC0-CT035
41	525.4	17.8	558	164	BE144266	BE144266 MR0-HT016
42	524.4	17.8	528	139	BE709658	BE709658 IL3-HT061
43	520.8	17.7	565	164	BE144270	BE144270 MR0-HT016
44	520.6	17.7	701	150	BF537398	BF537398 602050035
45	520.4	17.6	738	140	BE787574	BE787574 601479463

ALIGNMENTS

RESULT	1	840 bp	EST	24-OCT-2000
AUI33372	AUI33372	NT2RP4 Homo sapiens	CDNA clone NT2RP4001941 5', mRNA	
LOCUS	AUI33372	840 bp	mRNA	
DEFINITION	AUI33372	NT2RP4 Homo sapiens	CDNA clone NT2RP4001941 5', mRNA	
ACCESSION	AUI33372			
VERSION	AUI33372			
KEYWORDS	AUI33372.1	GI:10993911		
SOURCE	EST.			
ORGANISM	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 840)			
AUTHORS	Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y., Saito, K., Yamamoto, J., Nishikawa, T., Nakamura, Y., Nagai, T., Sugano, S., Masuh, Y. and Isogai, T.			
TITLE	HRI human cDNA project (Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y.,			

JOURNAL COMMENT

Saito, K., Yamamoto, J., Nishikawa, T., Nakamura, Y., Nagai, T., Sugano, S., Masuh, Y., Isogai, T.)  
Unpublished (2000)  
Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; CDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES

source  
1. 840  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP4001941"  
/clone\_lib="NT2RP4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."  
BASE COUNT 212 a 215 c 207 g 203 t 3 others  
ORIGIN

Query Match 28.0%; Score 827; DB 108; Length 840;  
Best Local Similarity 99.0%; Pred. No. 3.8e-212;  
Matches 830; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY	128	cagcagagagagcaagatgctacgagcagggggtgctctgggctgacccgggaagtcg	187
DB	3	CAGCAGAGAGGCAAGATGCTACGGCAGGGGGTGGCTGTGGGGCTGACCCGGGAAGTCG	62
QY	188	ggtctcctcgccctctgctctctctgctgctcctactgacaggttctgcaagggaactca	247
DB	63	GGTCTCCTCGCCCTTGTCTTCTGCTGCTCTACTACGAGGTTGTGACAGGGAACCTCA	122
QY	248	gtgagagagagatataatcccttaataaacagctccctgttgcgctgctcaac	307
DB	123	GTGAGAGAGGAAGATATATATCCCTTAATAAACAGCTCCCTGTGCTGCTCAAC	182
QY	308	gccatcatcagattgctgctgctcctcaatagtgagacacaggggtatcacagctc	367
DB	183	GCCATCATCAGATTGCTGCTGCTCAATTAGTGAGACACAGGGGTATCACAGCTA	242
QY	368	gtagagagagagagagacacacagtggttactgacgagcccaaccccttaccatg	427
DB	243	GTAGAGAGAGAGAGACCTACAGTGGGTATTGACTGATGCCCCCAACCCCTTACATG	302
QY	428	gttctgctgagagagcaagcatttaccagggtttaatgagagagctgaaggaggaacc	487
DB	303	GTTCTGCTGAGAGAGCAAGCATTTTACAGAGGATTTAATGAGAGAGCTGAAGGAGAAC	362
QY	488	agccgaattgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct	547
DB	363	AGCCGAATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	422
QY	548	cctagtgtagcagtgcccaaatgagtggttgggttacttcaatctctatggcagag	607
DB	423	CCTAGTGTAAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	482
QY	608	tttgctcctcagagagaaatacagtggaattcgctgggcaatggttggctatgaagac	667
DB	483	TTTGCTCCTCAGAGAGAAATACAGTGAATTCGCTGGGCAATGCTTGGCTATGAGAC	542
QY	668	tttagttcccatcttctcttgaagatgaataaagaaacaaagtcatacaagcagtc	727
DB	543	TTTAGTTCCCATCTTCTTCTTGAAGATGAATAAAGCAAGTCAATCAAGCAGTGC	602
QY	728	tatcaagatcacacactgagtcagatgctcagccacacacttccactatgtgcatg	787

|||||  
Db 603 TATCAAGATGACAACTGAGTGCAGATGGCTCAGACCAACCTTCCACTATGTGCCATG 662  
QY 788 cagctctttcaacacatgcatgctgttcacgacactgccaactgcatgcygcagctcc 847  
Db 663 CAGCTCTTTTACACATGCATGCTGTCTATCAGACACTGCCACTGTCATGGCGCAGCTTC 722  
QY 848 atccaaagcaccttcagcatcaacccaagaatcgtctgtgacccctgtctgattacaat 907  
Db 723 ATCCAAAGCACTTCAGCATCAACCCGAGAAATCGTCTGTGACCCCTGTCTGAATACCAT 782  
QY 908 gtgtgagcatgtctaaagcctataatacaactggagacattaaagcctgacagcaggy 965  
Db 783 GTGTGAGCATGCTAAAGCCTATATAATACCACCTGGACATTTNAGCCTGACNACANGG 840  
RESULT 2  
BG489641 853 bp mRNA EST 27-MAR-2001  
LOCUS 602518064F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4636670 5',  
DEFINITION mRNA sequence.  
ACCESSION BG489641  
VERSION BG489641.1 GI:13451151  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE NIH-MGC http://mgi.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: DCTD/DTP/Gazdar  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLCM1395 row: p column: 15  
High quality sequence stop: 853.  
Location/Qualifiers  
1. 853  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4636670"  
/clone\_lib="NIH\_MGC\_18"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC library."  
BASE COUNT 202 a 237 c 211 g 203 t  
ORIGIN  
Query Match 27.0%; Score 797; DB 154; Length 853;  
Best Local Similarity 97.4%; Pred. No. 4.8e-204;  
Matches 831; Conservative 0; Mismatches 20; Indels 2; Gaps 2;  
QY 1324 ggagagatctctggccacattggagaagatggtgtgtgtcctgtcatcctcag 1383  
Db 2 GGAGATCTCTCTGGCCACATTTGAGAGAGTGTGTGTCTCTCTGTCTCATCTCTCAG 61  
QY 1384 gagccaatcagtcaccagcctctccacacatctccctgcagcagattctcagctcg 1443  
|||||

Db 62 GAGGCCAATCAGTCCCAAGCCTCTCCACCACTCTCCCTGACAGCATTTCTCGAGCTCG 121  
QY 1444 aaacatctctgagctgttctctgtgctgacacactctgtgcttccataacaataatacga 1503  
Db 122 AAACATCTCTGGCGTGTCTCTGGCTGACCACTCTGCTGCTTCCATTAACAATATTACCA 181  
QY 1504 gagtattacgacacactgtgagaacattatgtgactataccgaatggtgagccctga 1563  
Db 182 GAGTATTACGACACTGCTGAGAACATTAATGTGAGCTATCCCGAATGGCTGAGCCCTGA 241  
QY 1564 agaggacctgaacttgtlaacagacactgccaagggccctgcaagatgtgccaagctgtgt 1623  
Db 242 AGAGGACCTGAACCTTTGTAAACAGACACTGCCAAGCCCTGGCAGATGTGGCCACGCTGCT 301  
QY 1624 ggagcgtgctctgtatgagcttgcagggaggaaccaacttcagcgacacagttcagctga 1683  
Db 302 GGGACGTGCTCTGTATGAGCTTGCAGAGAGGAACCAACTTCAGCGACACAGTTCAGGCTGA 361  
QY 1684 tccccaacaggttaccccgctgctctatgtgttctctgattaaagccaactcatgtgt 1743  
Db 362 TCCCAACAGCGTTACCCCGCCTGCTCTATGGGTTCTTGATTAAGCCCAACAACATCATGTT 421  
QY 1744 ccagctatctctcagcagcagacctaagttcctacttctgtgagcggcctcttcaacatta 1803  
Db 422 CCAGTCTATCTCAGGACGAGACCTTAAGTCTTACTTGGGTGACGGGCTCTTCAACATTA 481  
QY 1804 catcgtgtctccagcccccacccaacacttatgtgtacagtatgtccttgacaaattt 1863  
Db 482 CATCGCTGTCTCCAGCCCCACCAACCACTTATGTGTGACAGTATGCCCTTGCCAAATT 541  
QY 1864 gactggcagatgtgtcaacctcaaccagagagcagtgccagagatccaagtaagttccaaag 1923  
Db 542 GACTGGCAGAGTGTCAACCTCAACCCGAGAGAGCAGTGCCAGATCCAAAGTCCCAAG 601  
QY 1924 tgaacaacagatctgtatgactactcatgtgtccagggcccttgatcttaagagac 1983  
Db 602 TGAACAACAGGATCTGTATGACTCATGTGGTCCAGGCGCCTTGCACTTCAATGAGAC 661  
QY 1984 ggaacgactcccccggtgtgtgtcttacttcagcagatagccagggcctgtctcctgc 2043  
Db 662 GGACCGACTCCCGGCTGTGTGCG-TCTACTGCACGATTAAGCAGGTCTGTCTCTGC 720  
QY 2044 ctltgaactgagtcagtgagctctactgaataactctacatgactgactgagagccgtggaa 2103  
Db 721 CCTGAACCTGAGTCAAGTGAAGCTTACTGATATCTCTACATGAGACTGAGAGCCGTGAA 780  
QY 2104 agatatacgtgccccgatatcttcacatgcgcagcaagaagccttgatgtgtaaccctg-p 2162  
Db 781 AGGATATCCGTCGCGGATATTTCTCATCGCAGCAAGAAAGAGCTGAGTTGATCACTCTGAA 840  
QY 2163 cagtgggcttcgg 2175  
Db 841 CAGTGGGCTTCGG 853  
RESULT 3  
AUI32380 838 bp mRNA EST 24-OCT-2000  
LOCUS AUI32380 NT2RP3 Homo sapiens cDNA clone NT2RP3004312 5', mRNA  
DEFINITION sequence.  
ACCESSION AUI32380  
VERSION AUI32380.1 GI:10992734  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,  
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and  
Isogai,T.  
HRI human cDNA project  
JOURNAL Unpublished (2000)

**COMMENT**

Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: [genomics@hri.co.jp](mailto:genomics@hri.co.jp)

HRI human cDNA project; 5' - & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

## FEATURES

**Source**

```

location/Qualifiers
1..838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3004312"
/clone_lib="NT2RP3"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

```

BASE COUNT	221 a	207 c	202 g	206 t	2 others
ORIGIN					

## ORIGIN

Query Match	26.0%;	Score 768.2;	DB 108;	Length 838;
Best Local Similarity	98.8%;	Pred. No. 2.9e-196;		
Matches 815; Conservative	0;	Mismatches 5;	Indels 5;	Gaps 4;

OY	132	agaagaagcaagatggtcctacgcaagggtgtgctctggtgctgacccgggaagtcggtc	191
Db	17	AGAGAGGCCAAGATGGCTACGGCAGGGGGTGGCTCTGGGGCTGACCCGGGAAGTCGGGGTC	76
OY	192	tccttcgacctctgtctctctgcgtccctactagcaggtttgtgcaggggaaactcagtyg	251
Db	77	TCCTTCGCTTCTGTCTTCTTCTGCGCTCTACTAGCAGGTTTGTGCAGGGGAACCTCAGTGG	136
OY	252	agaagaagatatatacccccttaataaaacagctccctgtgttcgctgctcaacgcga	311
Db	137	AGAGGAAGATATATATATCCCTTAATATAAACAGACTCCCTGTGTCCCTTCACAGCCA	196
OY	312	ctcatcagattgctgctccagctctcaattagtgagacacaggggtatccacgtagtag	371
Db	197	CTCATCAGATTGGCTGCCAGTCTTCAATTAGTGAGACACAGGGGTTATCCACCTAGTAG	256
OY	372	agaagaagaggaacctacagtggtatagctgcatggtgcccccaacccccctacatggtc	431
Db	257	AGAAAGAGGAGGACCTACAGTGGGTATTGATGATGCGCCCCCAACCCCTTACATGGTTC	316
OY	432	tgcctgagagcaagcatttaccaggatattaatgagaaagctgaaagaggaaccagcc	491
Db	317	TGCTGAGAGAGCAACATTTTACCAGGATTTAATGAGAAAGCTGAAAGGAGAACCAAGCC	376
OY	492	gaattgctgtctctgcagtgctccttgaccaagccccagctcctcagcgtctctccta	551
Db	377	GAATTGCTGCTCTTCAGAGTGCCTTGACCAAGCCCACTGCTGCCTCAGGCTTCTCCTA	436
OY	552	gtgtacagtgcccaaatgatgtggtttgtgtttactccaattcctatggtccagagtttg	611
Db	437	GTGTACAGTGCCCAATGATGGGTTTGGTGTATTACCAATTCCATGAGCCAGAGTTTG	496
OY	612	ctcactgcagagaaatatacagtggaattcgctgggcaatggtttggtatgaagacttta	671
Db	497	CTCACTGCAGAGAAATACAGTGGGAATTCGCTGGGCAATGGTTTGGCTTATGAAGACTTTA	556
OY	672	gtttcccccattctctctctgaagatga <sup>1</sup> aatgnaaaccaaagtcatcaagcagtgctatc	731
Db	557	GTTTCCCATCTTCTTCTTTGAAGATGAAAATGAAACCAAGTCATCAAGCAGTGTATC	616
OY	732	aagatcaacaacctgagtcagaatggtcagcaccacaccttcccactatgtgcacatgcagc	791
Db	617	AAGATCAACAACCTGAGTCAGATGCGCTCAGCACCAACCTT-CCACTATGTGCCATGCAGC	675

[illegible]

RESULT 4  
AI377767/c  
LOCUS  
DEFINITION  
AI377767 784 bp mRNA EST 28-MAR-1999  
te62b05.x1 Soares.NFL\_T-GBC\_S1 Homo sapiens CDNA clone  
IMAGE:2091249 3' similar to TR:Q92542 Q92542 MYELOBLAST K1AA0  
mRNA sequence.  
ACCESSION  
AI377767  
VERSION  
AI377767.1 GI:4187620  
KEYWORDS  
EST.  
SOURCE  
human.

JOURNAL COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 450    Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 462.
FEATURES	Location/Qualifiers
source	1. .784

## FEATURES

**Source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2091249"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCL-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

```

BASE COUNT	238 a	141 c	232 g	172 t	1 others
ORIGIN					

## ORIGIN

Query Match	25.68;	Score 753.8;	DB 19,	Length 784;	1;
Best Local Similarity	98.38;	Pred. No. 2.2e-192;			
Matches 769;	Conservative	2;	Mismatches 10;	Indels 1;	Gaps 1;

QY	2168	ggtctgcgcatacctcatcttctccctcatcgtaacctactgcataatgccaaagctgat	2227
Db	784	GGGTTCGGCATCCTACTTCTCCCTCATCGTAACCTACTGCATCAATGCCANAGCTGAT	725
QY	2228	gtcccttttcatttgtctccccggaggacagagctgtgtcatactagsagaccsagaatt	2287
		:       :	
Db	724	GTCCTTTTCATTGCTCCCCGGAGAGCCAGAGAGCTGTGTCACTACTGAGGAGAACCCAGAGCTT	665







Db 84 CCTAGTTACCCACCCTAA-TTGCCCTTCAGGACCCTTCTACATTTCGCCGCCGCCCTCGG 26  
QY 2615 accctctctgtctcctcacc 2635  
1 | 1 | 1 | 1 | 1 | 1 |  
Db 25 CCGATTACACCGCGCTCGACC 5

RESULT 6  
BF339787 914 bp mRNA EST 22-NOV-2000  
LOCUS 602034987F1 NCI\_CGAP\_Brn64 Homo sapiens cDNA clone IMAGE:4183101  
DEFINITION 5', mRNA sequence.  
ACCESSION BF339787  
VERSION BF339787.1 GI:11286246  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 914)  
NIH-MGC http://mgs.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: David N. Louis, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLM9498 row: m column: 22  
High quality sequence stop: 670.  
Location/Qualifiers

FEATURES  
source 1..914  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4183101"  
/clone\_lib="NCI\_CGAP\_Brn64"  
/tissue\_type="glioblastoma with EGFR amplification"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: brain; Vector: PCMV-SPORE6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.57 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 197 a 268 c 226 g 223 t  
ORIGIN

Query Match 23.0%; Score 677.8; DB 147; Length 914;  
Best Local Similarity 97.0%; Pred. No. 8e-172;  
Matches 752; Conservative 2; Mismatches 14; Indels 7; Gaps 6;

QY 1570 cctgaacttgttaacagacactgccaagccctggcagatgtgccaagtgctggagc 1629  
|||||  
Db 1 CCTGAACCTTGTAAACAGACACTGCCAGGCCCTGGCAGATGTGCCACGCTGGGACG 60

QY 1630 tgcctgtatgagcttgcaagaggaaccaacttcagcgacacagttcaggtgatccca 1689  
|||||  
Db 61 TGCTCTGTATGAGCTTGCAAGAGGAGCAACTTCAGCGACACAGTTCAGGCTGATCCCA 120

QY 1690 aacggttaaccgctgtctataggttctgtatgaagccaacaactcatgttccagtc 1749  
|||||  
Db 121 AAGGTTACCCGCTGCTATGGGTCTGATTAAGCCACAACTCATGTCCAGTC 180

QY 1750 tatcctcagcaggaacctaaagtcctacttgggtgagcggccttcaacattacatgc 1809  
|||||  
Db 181 TATCCTCAGGAGGAGCACTAAGGCTCTACTTGGGTGACGGGCTCTTCAACATTACATCGC 240

QY 1810 tgcctcagcccccaacaacacactatgtttgtacagtatgcttggcaaatgtactg 1869  
|||||  
Db 241 TGCTCTCAGGCCCCCACCACCACTTATGTGTACAGTATGCCCTTGGCAAAATTGACTGG 300

QY 1870 caacgtgtcaaccctcaaccagagcagtgccagatcccaagtaagtcaccaagtgaap 1929  
|||||  
Db 301 CACAGTGTCAACCTCACCCGAGAGCAGTGCAGATCCAAAGTCCCAAGTGA 360

QY 1930 caagatctgtatgactcatgtgtccagggcccttgcattctaatgagacgacg 1989  
|||||  
Db 361 CAAGATCTGTATGAGTACTCATGGGTCCA-GGCCCTTTCATTCTAATGAGACGGACCG 419

QY 1990 actcccccggtgtgtgttctactgcacgattagccaggccctgtctccttga 2049  
|||||  
Db 420 ACTCCCCCGGTGTGTGCTTCTACTGCACGATFAGCCAGGCGCTTGTCTCTGCTTGA 479

QY 2050 actgagtcagtgagctctactgaatactctacatgactgagagccgctggaagatat 2109  
|||||  
Db 480 ACTGAGTCAGTGAGGCTCTACTGAATACTCTACATGAGACTGAGAGCCGCTGGAAGATAT 539

QY 2110 ccgtgccccgatatattctcctcgcagcaagagcttgagttgatccacctgacagtg 2169  
|||||  
Db 540 CCGTGCCCGGATATTCTCTATTCGCCCAAGCAAGAGCTTGATCACCCTGACAGTGGS 599

QY 2170 ctgcgcatccctcatctctcctcatcgtcactgcatcatcaatgccaagtgtatgt 2229  
|||||  
Db 600 CTTGGCATCTCATCTTCTCCCTCATCTCATCTACTG-ATCAATGCCAAGCTGATGT 658

QY 2230 cctttcattgtctcccgagag--ccaagagctgtgtacac-tgagsagaccsagct 2286  
|||||  
Db 659 CCTTTTCATTCTCCCGGAGGAGGAGGAGGCTGGGTCTACTTGGGAGGAGAACCCAGGT 718

QY 2287 ttctctgcagctc-agcagttcactt-cttagagcatctgtcccactgagacac 2339  
|||||  
Db 719 TTTCTTGAGAGCTCAAGCAGTTCACTTCCCTAGAGCATTTGTCCCTTGGGACACAC 773

RESULT 7  
BG568503 696 bp mRNA EST 10-APR-2001  
LOCUS 602587478F1 NIH\_MGC\_76 Homo sapiens cDNA clone IMAGE:4716326 5',  
DEFINITION mRNA sequence.  
ACCESSION BG568503  
VERSION BG568503.1 GI:13576156  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 696)  
NIH-MGC http://mgs.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: CLONETECH Laboratories, Inc.  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLM1564 row: k column: 19  
High quality sequence stop: 696.  
Location/Qualifiers

FEATURES  
source 1..696  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4716426"  
/clone\_lib="NIH\_MGC\_76"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgcctggcgc); Site\_2: SfiI (ggcattatggcgc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGGCCATATGAGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGGCCGAGGCGGCGGACATG-dT(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."

Query Match	22.6%;	Score 666.2;	DB 155;	Length 696;
Best Local Similarity	97.9%;	Pred. No. 1e-168;		
Matches 684; Conservative	2;	Mismatches 10;	Indels 3;	Gaps 1;

QY	1974	ctaattgaaagcagccgactccccggtgtgtgcgttctactgcacgattagccaagccct	2033
Db	1	CTAATGAGACGACCGACTCCCCCGGTGTGTGCGTTCTACTGACAGATTAGCCAGGGCCT	60
QY	2034	tgtctcctgccttgaactgagtcagtygagcttactgaatactctacatgactgga	2093
Db	61	TGTCTCTGCTCTTGAACGTAGTCAGTGGAGCTCTACTGAATACTCTACATGGACTGAGA	120
QY	2094	gccgcctgaaaagatatccgtgcccgcgatatctcatalcgcgaagcaagagcttgattga	2153
Db	121	GCCGCTGGAAGATATCCGTTGCCGGATATTTCATCGCCACCAAGAGCTTGAGTTGA	180
QY	2154	tcaccctgacagtgygcttcggcatccctcatctctccctcatcgtcaacctactgcat	2213
Db	181	TCACCCCTGACAGTGGGCTTCGGCATCTCATCTTCTCCTCATCGTCACCTACTGCATCA	240
QY	2214	atgccaaaagctgatgtcccttttcaattgtccccgcggaagcagagctgtgtcatactgag	2273
Db	241	ATGCCAAAGCTGATGTCCCTTTTCATTGCTCCCCGGAGCAGGAGCTGTGTCACTAGAG	300
QY	2274	saggaccscagctttctctgcccagctcagcagttcaccttccctctagagcatctgtcccactg	2333
Db	301	GAGGACCCCGAGCTTTCTTGCCAGCTCAGCAGTTCACTTCTTAGAGCATCTGTCCCACTG	360
QY	2334	ggacacaaccactaatttgtcacttgaacctccctcgtgctgtctcagattgagattaac	2393
Db	361	GGACACAACCACTAATTTGTCACTGGAACCTCCCTGGGCCGTCTCAGATTGGGATTAAAC	420
QY	2394	ataaaaagatgnaactatccaaaaagagacaggagagaaataataatgtccctccctcct	2453
Db	421	ATAAAAAAGTGAACTAATCCAAAAAGACAGGAGAGAAATAATAATTGCTCCCTTCTCT	480
QY	2454	ccgcctccctttcccatcacccttccccaattccctcttccctctctctactcatgccaagat	2513
Db	481	CCGCTCCCCCTTCCCATCAACCCCTTCCCATTTTCTCTTCTCTCTACTCATGCCCAGAT	540
QY	2514	tttggattacaataagaagctctctgtcctcgtttaactccctagttaccaccctaata	2573
Db	541	TTTGGATTACAATAAGAAAGCTTCTGTCTCCTGTTAACTCCCTAGTTACCACGCTAAT	600
QY	2574	ttgcctctcagacaccttctacttttctctccctgcgccctgtacacctctctctgtcctcac	2633
Db	601	TTGCCCTTCAGGACCTTCTACTTTTTCCTTCCTGCCCTGTACCTCTCTCTGCTCTCTCAC	660
QY	2634	ccccaccctgtaccaccagccaccttctctgactggaag	2672
Db	661	CCACACCC--TGTACCAAGCAACTTCTCTGACTGGGAAGG	696

RESULT	8
BF340237	
LOCUS	
DEFINITION	BF340237 910 bp mRNA
	602036827F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4184814
	5', mRNA sequence.
ACCESSION	BF340237
VERSION	BF340237.1 GI:11286699
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 910)	NIH-MGC	<a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	National Institutes of Health, Mammalian Gene Collection (MGC), Unpublished (1999)	Contact: Robert Strausberg, ph.D.

Tissue Procurement: David N. Louis, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LHAM9503 row: e column: 07  
High quality sequence stop: 711.

FEATURES  
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1="IMAGE:4184814"
/clone_1b="NCI_CGAP_Brn64"
/tissue_type="glioblastoma with EGFR amplification"
/lab_host="DH10B (T1 phage-resistant)"
/name="Organ: brain; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.57 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      210 a      271 c      179 g      249 t
ORIGIN           1 others

```

Query Match	22.48;	Score 661.4;	DB 147;	Length 910;
Best Local Similarity	95.48;	Pred. No. 2.2e-167;		
Matches 744; Conservative	2;	Mismatches 24;	Indels 10;	Gaps 6

QY	2151	tgatcaacctgacagatgtgggtctgggcatcctcatctctccctcatcgtaacctactgca	2210
Db	1	TGATCACCCCTGACAGTGGGCTTGGGCATCCTCATCTTCTCCCTCATCGTCACCTACTGCA	60
QY	2211	tcaatgcccaaaagctgatgttcctttcattgtctcccgaggaccaagagctgtgtcatact	2270
Db	61	TCAATGCCCAAAGCTGATGTCTCTTTTCATTGCTCTCCCGGAGCAGGAGCTGTGTCACTACT	120
QY	2271	gagsaaggaccscagcttctcttgcacgctcagcagttcaattccttagagcatctgtccca	2330
Db	121	GAGSAGGACCCCGAGCTTTCTTGTGCCAGCTCAGCAGTTCACTTCTTAGAGCATCTGTCCCA	180
QY	2331	ctgagacacaaccactaatctgttcaactgagaaacctcccttgggcctgtctcagattggatt	2390
Db	181	CTGGGACACACAACCACCTAATTTGTCACTGGAACCTCCCTGGGCCCTGTCTCAGATTGGGATT	240
QY	2391	aacataaaaagagtgyaactatccaaaagagacagggagaaataaataattgtccttcctt	2450
Db	241	AACATAAAAAGAGTGAACACTATCCCAAAGAGACAGGAGAAATAATAATTGCTTCCTT	300
QY	2451	cctccgctcccttcccatcaccccttccccatttcccttcccttctctactcatgcca	2510
Db	301	CCTCCGCTCCCTTTCCCATCACCCCTTCCCCAATTCTCTTCTCTACTCATGCCA	360
QY	2511	gatttgggattacaataagaagctctgtctcctgtttaactccctagttaccaccct	2570
Db	361	GATTTGGGATTACAATAAGAAGCTTCTGTGCTCCTGTTAACTCCCTAGTTACCCACCCCT	420
QY	2571	aatttggccttcagagacccttctacttttcttccctcctgcgcctgtacactctctctgcct	2630
Db	421	AATTTGCCCTTCAGGACCCCTTCTACTTTTCTCTCCCTGCGCCCTGTACCTCTCTGCTCCCT	480
QY	2631	cacccccaccctgttacccagccacactctctgtactgtggaagacataaaagtttaattgt	2690
Db	481	CACCCCCACCCCTGTATCCCCAGCCACCTTCTCTGACTGGGAAGACATAAAGGTTTAATGT	540
QY	2691	cagggtcaactacattgagccctcctgaggaacaggggcatctctggtgctgagcctactgtt	2750

[illegible]

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLCM338 row: 1 column: 23  
 High quality sequence stop: 610.

FEATURES	Location/Qualifiers
source	1. .807

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3638182"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

Query Match	22.1%;	Score 652.8;	DB 167;	Length 807;
Best Local Similarity	95.8%;	Pred. No. 4.4e-165;		
Matches 692;	Conservative	0;	Mismatches 27;	Indels 3;
				Gaps 22;

QY	938	actggyacattaaagccctgacgacagaggttgtgtgtcgtccaccgccgctgatatgct	997
Db	1	ACTGGACATTAAAGCCCTGACGACAGAGGTTGTGTGTGCTGCCACCCGGCTGATAGTGT	60
QY	998	tccttttctgnaatgtggtcccccaggggctgaaagcgcagtgtgctcctttgtcaccag	1057
Db	61	TCCTTTTCTGGAATGTGCCCCCAGGGGCTGAAAGCGcAGTGGCTTCCCTTTGTACCCAG	120

QY	1058	ctgctgctgctgaagctlltgcaaaagcaccctgatagtgacaccctgccccgaatgct	1117
Db	121	CTGGCTGCTGCTGAAGCTTTGCAAAAGGCACCTGATGTGACCACCCTGCCCCGAATGTC	180
QY	1118	atgtttgtctctcttcaagggaacatttgactacattggcagctcgagatggtctac	1177
Db	181	ATGTTTGTCTCTTTCAGGGGAAACTTTGACTACATTGGCAGCTCGAGGATGGTCTAC	240
QY	1178	gatatgagaagggcaagttccocgtgcagttagaagaatgtgactcattgtgagctg	1237
Db	241	GATATGGAGAGAGGGCAGACTTTCCCGTGCAGTTAGAGAAFTGTGACTCATTTGTGGAGCTG	300
QY	1238	ggacaggtgyccttaagaaacttcattagagcttggatgacacacagatccctgtttctcag	1297
Db	301	GGACAGGTGGCCTTAAGAACTTCATTAGAGCTTTGGATGCACACAGATCCTGTTCCTCAG	360
QY	1298	aaaaatgagtctgtacggaaccaggttgagagatctctgycacacatttgaagaagttgt	1357
Db	361	AAAAATGAGTCTGTACGGAAACCAGGTGGAGATCCTCGGCCACATTGGAGAGAGTGGT	420
QY	1358	gctggtgtccctgctgtcatcctcaagagagccaaatcagtcaccagcctctccaccatct	1417
Db	421	GCTGGTGTCCCTGCTGTCAATCCTCAGAGGGCCAAATCAGTCCCAAGCTCTCCACCATCT	480
QY	1418	tcctctgacagcgaattctctcagagctcgaaacatctctgycgttgttctgctgacacatct	1477
Db	481	TCCCTGCAGCGATTCTTCGAGCTCGAAACATCTCTGGCGTGTCTTGCGTGACCACTCT	540
QY	1478	ggtgccttccataacaataattaccagagtaatttaagacacactgctgagaacatlaatgtg	1537
Db	541	GGTGCCCTTCATTAACAATAATTACCAGAGTATTTACGACACTGCTGAGAACATTAATGTS	600
QY	1538	agctatcccgaaatgctgagccct--gaagaggaactgaaacttgtlaacag-acactgcc	1594
Db	601	AGCTATCCCGAAATGCTGAAGCCTTGAAAAGGGAGACTGTACTTTGTTACAGAACCCTGCC	660
QY	1595	aagggccctgacagatgttgccacaggtgctgagacgtgtctgtatgagcttgcagagga	1654
Db	661	AAGGCCCTGGCAATGTGTGGCCCGGGGCTGGGACCTGCTCTGTGTAGCTTGCGGGGGCA	720
QY	1655	ac 1656	
Db	721	CC 722	

RESULT	10
BF980176/c	
LOCUS	
DEFINITION	BF980176 950 bp mRNA
	602288192T1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4373831 3'
	mRNA sequence.
ACCESSION	BF980176
VERSION	BF980176.1 GI:12347391
KEYWORDS	EST.
SOURCE	human.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 950)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraiki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10036 row: h column: 24









REFERENCE 1 (bases 1 to 934)  
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: David N. Louis, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LAM9789 row: e column: 19  
High quality sequence stop: 657.

## FEATURES

source  
1. 934  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:4214778"  
/tissue\_type="anaplastic oligodendroglioma with lp/19q  
loss"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: brain; Vector: PCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.3 kb. Constructed by Life  
Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 222 a 274 c 212 g 226 t  
ORIGIN

Query Match 21.3%; Score 627; DB 150; Length 934;  
Best Local Similarity 93.8%; Pred. No. 4.3e-158;  
Matches 727; Conservative 2; Mismatches 37; Indels 9; Gaps 7;

```
QY 1713 ggttcctgattaagcccaactcatgttccagctctacccagcaggaactaagt 1772
    |||||||
Db 1 GGTTCCTGATTAAAGCCCAACTCATGTGCTCCAGTCTATCCTCAGGACGACCTAAGT 60
QY 1773 cctacttggtgacggcctcttcaacattacatgctgtctccagccccaacacaa 1832
    |||||||
Db 61 CCTACTTGGGTGACGGGCTCTTCAACATTACATCGCTGCTCCAGCCCAACACACA 120
QY 1833 ctatgtgtacagatgcttggcaattgaactggcacagtgtgaacctcaccggag 1892
    |||||||
Db 121 CTTATG-TGTACAGTATGCTT--GGCAATTTGACTGGCACAGTGTGTAACCTCA 177
QY 1893 agcagtgccagatccaagtaaaagtcaccaagtgaaacaagatctgtatgactcat 1952
    |||||||
Db 178 AGCAGTGCCAGGATCCAAAGTCCCAAGTGAAGAAACAAGGATCTGTATGAGTACTCAT 237
QY 1953 ggggtcaggcccttgcattcaatgagcagcagccgactccccgggtgtgtcgtcta 2012
    |||||||
Db 238 GGGTCCA-GGCCCTTGTGACTTCTAATGAGACGACCGACTCCCCGGGTGTGTGCTTCTA 296
QY 2013 ctgacgattagcaggcctgtctcctgcttgaactgagtcagtgagctcactg 2072
    |||||||
Db 297 CTGACGATTAGCCA-GGCCCTTGTCTCTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGA 355
QY 2073 aatactctacatgactgagagccgctgggaaagatatccgtgccggatattctcatcg 2132
    |||||||
Db 356 AATACTCTACATGAGCTGAGAGCCGCTGGAAGAAGATATCCGTGCCGGATATTCTCATCG 415
QY 2133 ccagcaagagcttgagttgatacccttgacagtggtggtcgcgacatctcctcctcc 2192
    |||||||
Db 416 CCAGCAAGAGCTTGAGTGTGATCACCTTGACAGTGGGCTTGGCATCTCTCATCTTCC 475
QY 2193 tcatcgtcactactgcatcaatgccaagctgattccttctcattgtccccggagc 2252
    |||||||
Db 476 TCATCGTCACTACTGATCAATGCAAGCTGATGTCTTTTCATTGCTCCCCGGGAGC 535
```

```
QY 2253 caagagctgtgtacatcagcagcagccagcttcttccagctcagcagcttcaact 2312
    |||||||
Db 536 CAGGAGCTGTGTACTACTGAGGAGGAGCCAGCTTTCTTCCAGCTCAGTCACTACTT 595
QY 2313 cctagagcatctgtcccaactgggacacacacactaattgtcactggaacctccctgggc 2372
    |||||||
Db 596 CCTAGAGCATCTGTCCCACTGGGACACACACTAA-TTGTACTGTGAACTCTCCCTGGGC 654
QY 2373 ctgtcagattgggattaacataaagagtggaactatcccaaaagagacagg-gagaag 2431
    |||||||
Db 655 CTGTCTCAGATGGGAATTACATAAA--GAGGGGACTATCCAGAGAGACGGGAGAAATA 712
QY 2432 taataaatgtccctccctccctccctccctccctccctccctccctccctccct 2485
    |||||||
Db 713 CATAAATGCTCCCTTCTCTGCTGCTCCCTTCCATCACTTCCATCTCTCTCTCT 767
```

RESULT 14  
BF569485 813 bp mRNA EST 12-DEC-2000  
LOCUS 602185971F1 NIH\_MGC\_45 Homo sapiens cDNA clone IMAGE:4310584 5',  
DEFINITION mRNA sequence.  
ACCESSION BF569485 GI:11642865  
VERSION BF569485  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 813)  
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: Linehan  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LCM1185 row: m column: 17  
High quality sequence stop: 742.

## FEATURES

source  
1. 813  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:4310584"  
/tissue\_type="renal carcinoma (ascites)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pORF7; Site\_1: XhoI; Site\_2:  
EcoRI; CDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH-MGC Library."

BASE COUNT 191 a 228 c 161 g 233 t  
ORIGIN

Query Match 21.1%; Score 623; DB 150; Length 813;  
Best Local Similarity 97.5%; Pred. No. 5e-157;  
Matches 715; Conservative 2; Mismatches 7; Indels 9; Gaps 8;

```
QY 2216 gccaaagctgattccttctcattgtctcccgaggagcagagctgtgtcactagagga 2275
    |||||||
Db 1 GCCAAAGCTGATGCTCTTTTCATTTGCTCCCGGAGGACGAGAGCTGTGTCACTAGAGA 60
QY 2276 ggaaccagcttctcttgccagctcagcagcttctcctagagcatctgtccactggg 2335
```

Db		61	GGACCCAGCTTTTCTTGCCAGCTCAGCAGTTCACTTCCTAGAGCATCTGTCCACATGGG	120
OY	2336	acacaaccactaatltgtcactggaacctccctgggcctgtctcagattggaattaacat		2395
Db	121	ACACAACCACATAA-TTGTCACTGGAACTCCCTGGGCCCTGTCTCAGATTGGGATTAACAT		179
OY	2396	aaagaagtgaactataccaagaagacaggagaagaataataaatgtccctccctcc		2455
Db	180	AAAAGAGTGAACACTATCCAAAAGAGACAGGAGAATAATAA-TGCCTCCCTTCTCC		238
OY	2456	gtccccctttcccatacaaccccttccccatttcctctcctctcactcatgccagatt		2515
Db	239	GCTCCCCCTTTCCCATCACCCCCTTCCCATTTCTCTCTCTACTCATGCGAGA-TT		297
OY	2516	tgggattacaataagaagctctgtcctgtttaactccctaagttaaccaaccataatt		2575
Db	298	TGGGATTACAATAAGAAGCTTC-TGCTCTGTTTAACTCCCTAGTTACCAACCTAA-TT		355
OY	2576	gcccttgaggaccccttctacttttctcctcctcgccctgtaacctctctgtcctcaccc		2635
Db	356	GCCCTTCAGGACCCCTTCTACTTTTCTCTCCCTGCCCCGTACCTCTCTGCTCCTCACCC		415
OY	2636	ccaccctgtacccagccaccccttcctgactggygaagacataaaaagtttaattgacagg		2695
Db	416	CCACCCTGTACCCAGCCACCTTCTGACTGGGAAGACATAAAAAGTTTAATGTCAAGG		475
OY	2696	tcaaatcatattgagccctgagagacaggggcatctctggtgtagcctaactgtccct		2755
Db	476	TCAAATCATATTGAGCCCCCTGAGGACAGGGGCATCTCTGGGCTGAGCCTACTGTCTCCTT		535
OY	2756	cccaactgtccttcttccagggccctcagaatgycacattaggtggtgctgtcgcggtg		2815
Db	536	CCCACCTGTCTTCTTCCAGGCCCCCTCAGATGGCACATTAGGGTGGCGCTGCGGGTGGG		595
OY	2816	tatcccacctccagccacagatgctcagtlgtactttlattaagctgaatatctattt		2875
Db	596	TATCCACCTCCAGCCACAGTGTCTAGTTGACTTTTTTAAAGCTGAATATCTATT		655
OY	2876	ttgttttgtcctttccttatcttttcttgaataataataataat-gagttcattaa		2934
Db	656	T--GTTTGTCTTTTCCCTTAATTC-TTTGTAATATATATATATATGAGTTGCATTAA		712
OY	2935	aatagattatccc	2947	
Db	713	AATTGAATTATCC	725	
RESULT	15			
LOCUS	BE788789	820 bp	mRNA	EST
DEFINITION	601475883F1 NIH_MGC_68 Homo sapiens CDNA clone IMAGE:3878780 5'			
ACCESSION	BE788789			
VERSION	BE788789.1	GI:10209976		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE	1 (bases 1 to 820)			
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/.			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-remail.nih.gov			
	Tissue Procurement: DCTD/DTP/Gazdar			
	CDNA Library Preparation: Life Technologies, Inc.			
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Incyte Genomics, Inc.			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	http://image.llnl.gov			

[illegible]

Thu Jun 28 10:23:28 2001

us-09-541-094-13.rst

Page 14

Search completed: June 27, 2001, 15:43:06  
Job time: 4080 sec







Qy	985	gctggatagtcgttcccttttcttggaatgtgccccagggtgaaagcgcaagtggcttc	1044
Db	841	gctggatagtcgttcccttttcttggaatgtgccccagggtgaaagcgcaagtggcttc	900
Qy	1045	ctttgtcaccagctggctgctgaagctttgcaaaaggcacctgatgtgaccacct	1104
Db	901	ctttgtcaccagctggctgctgaagctttgcaaaaggcacctgatgtgaccacct	960
Qy	1105	gccccgaatgtcatgtttgtctcttcaagggggaaacttttgactacattggcagctc	1164
Db	961	gccccgaatgtcatgtttgtctcttcaagggggaaacttttgactacattggcagctc	1020
Qy	1165	gagatggtctacgatatggagaaggcaagtctcccgtagttagagaatggtgactc	1224
Db	1021	gaggatgggtcagcatatggagaaggcaagtctcccgtagttagagaatggtgactc	1080
Qy	1225	atttgtgagctgggacaggtggccttaagaacttcattagagctttggatgcacacaga	1284
Db	1081	atttgtggagctgggacaggtggccttaagaacttcattagagctttggatgcacacaga	1140
Qy	1285	tcctgtttctcagaaaaatgagtctgtacggaaccaggtggaggatctcctggccacatt	1344
Db	1141	tcctgtttctcagaaaaatgagtctgtacggaaccaggtggaggatctcctggccacatt	1200
Qy	1345	ggagaagagtggtgctggtccctgctgcacatcctcaggaggccaaatcagtcaccgcc	1404
Db	1201	ggagaagagtggtgctggtccctgctgcacatcctcaggaggccaaatcagtcaccgcc	1260
Qy	1405	tctcccaccatctccctgcagcgattctctcgagctcgaaacatctctggcgttgtct	1464
Db	1261	tctcccaccatctccctgcagcgattctctcgagctcgaaacatctctggcgttgtct	1320
Qy	1465	ggctgaccactctggtgccttccataacaaatattaccagagtatttaacgacactgctga	1524
Db	1321	ggctgaccactctggtgccttccataacaaatattaccagagtatttaacgacactgctga	1380
Qy	1525	gaacattaatgtagctatcccgaatggctgagccctgaagaggacctgaactttgtaac	1584
Db	1381	gaacattaatgtagctatcccgaatggctgagccctgaagaggacctgaactttgtaac	1440
Qy	1585	agacactgccaaggccctggcagatgtggccacggtgctgggacgtgctctgtatgagct	1644
Db	1441	agacactgccaaggccctggcagatgtggccacggtgctgggacgtgctctgtatgagct	1500
Qy	1645	tgcaggaggaaaccaacttcagcgacacagttcaggctgatccccaaacggttacccgcct	1704
Db	1501	tgcaggaggaaaccaacttcagcgacacagttcaggctgatccccaaacggttacccgcct	1560
Qy	1705	gctctatgggttcctgattaaagccaacactcatggttccagtcctatcctcaggcagga	1764
Db	1561	gctctatgggttcctgattaaagccaacactcatggttccagtcctatcctcaggcagga	1620
Qy	1765	cctaagggtcctacttggtgacgggcctcttcaacattacatcgctgctccagccccac	1824
Db	1621	cctaagggtcctacttggtgacgggcctcttcaacattacatcgctgctccagccccac	1680
Qy	1825	caacaccacttatgttgtacagtatgccttggaatttgactggcacagtgggtcaacct	1884
Db	1681	caacaccacttatgttgtacagtatgccttggaatttgactggcacagtgggtcaacct	1740
Qy	1885	cacccgagagcagtgccaggaatccaaagtaagtcaccaagtgaacaaaggatctgtatga	1944
Db	1741	cacccgagagcagtgccaggaatccaaagtaagtcaccaagtgaacaaaggatctgtatga	1800
Qy	1945	gtactcatgggtccagggccctttgcatcttaatgagcgggacccgactccccgggtgtgt	2004
Db	1801	gfactcatgggtccagggccctttgcatcttaatgagcgggacccgactccccgggtgtgt	1860
Qy	2005	gcgttctactgcacgattagccagggccttgtctcctgcctttgaactgagtcagtgagg	2064
Db	1861	gcgttctactgcacgattagccagggccttgtctcctgcctttgaactgagtcagtgagg	1920

QY	2065	ctctactgaataactctacatggactgagagccgctggaagatatccgtgcccggatatt	2124
Db	1921	CTCTACTGAATACTCTACATGGACTGAGAGCCGCTGGAAGATATCCGTGCCGGATATT	1980
QY	2125	tctcatgccagcaagagccttgagttgatcacccctgacagtggtggttcggcatcctcat	2184
Db	1981	TCTCATCGCAGCAAGAGCCTTGAGTTGATCACCCCTGACAGTGGGCTTCGGCATCCTCAT	2040
QY	2185	cttctccctcatogtcacctactgcatcaatgccaaagctgatgtcctttttcattgctcc	2244
Db	2041	CTTCTCCCTCATCGTCACCTACTGCATCAATGCCAAAGCTGATGTCTCTTTCATGTCTCC	2100
QY	2245	ccgggagccaggagctgtgtcatactgagsgaggaccscagcttttcttgccagctcagca	2304
Db	2101	CCGGAGCCAGGAGCTGTGTATCTAGGAGGACCCAGCTTTTCTTGCCAGCTCAGCA	2160
QY	2305	gttcacttccctagagcatctgtcccactgaggacacaaaccactaatattgtcactggaacct	2364
Db	2161	GTTCACTTCCCTAGAGCATCTGTCCCACCTGGGACACAAACCACATAATTGTCACTGGAACCT	2220
QY	2365	ccctgggacctgtctcagattgggattaaacataaaagagtggaactatccaaaagagacag	2424
Db	2221	CCCTGGGCCCTGTCTCAGATTGGGATTAAACATAAAAGAGTGGAACTATCCAAAAGAGACAG	2280
QY	2425	ggagaaaataaaattgcctccctctccgctccctttcccatcaccccttcccccat	2484
Db	2281	GGAGAAATAAAATAAATTGCCTCCCTTCCCTCCGCTCCCTTTCCCATCACCCCTTCCCCAT	2340
QY	2485	tctctctctctctactcatgccagattttgggattacaaaatagaagcttcttctgctcc	2544
Db	2341	TTCTCTCTTCTTCTACTCATGCCAGATTTTGGGATTACAAATAGAAGCTTCTTGCTCC	2400
QY	2545	tgtttaactccctagttacccaccctaaattggcccttcaggacccttctactttttcctt	2604
Db	2401	TGTTTAACCTCCCTAGTTACCCACCCTAAATTGGCCCTTCAGGACCCCTTCTACTTTTTCCTT	2460
QY	2605	cctgccctgtacctctctctctcctcaccaccccccctgtacccagccaccttccctgac	2664
Db	2461	CCTGCCCTGTACCTCTCTCTGTCTCTCACCCCCACCCCTGTACCCAGCCACCTTCTCTGAC	2520
QY	2665	tgggaaggacataaaagggttaatgtcaggtgcacaaactacattgagccccctgaggacagg	2724
Db	2521	TGGGAAGGACATAAAAGGTTTAATGTCAAGGGTCAAACTACATTGAGCCCCCTGAGGACAGG	2580
QY	2725	ggcatctctgggctgagcctactgtctcctccactgtcctttctccaggccctcagat	2784
Db	2581	GGCATCTCTGGGCTGAGCCTACTGTCTCCTTCCCACCTGTCTCTTCTCCAGGCCCTCAGAT	2640
QY	2785	ggcacattagggtggcggtgctgcgggtgggtatcccacctccagccacacagtgtcagt	2844
Db	2641	GGCACATTAGGTGGCGGTGCTGCGGGTGGGTATCCACCTCCAGCCACACAGTGTCTAGT	2700
QY	2845	tgtactttttattaaagctgtaataatctatttttggtttttggctttttcctttattctttt	2904
Db	2701	TGTACTTTTATTAAAGCTGTAATATCTATTTTGTGTTTTTGTCTTTTCCCTTATTCTTTT	2760
QY	2905	tgtaaaatatatatataatgagtttccattaaaaatagattatcccac	2949
Db	2761	TGTAAATATATATATAATGAGTTTCATTAAATAGATTATCCCAC	2805
RESULT	3		
AF240469			
LOCUS	AF240469	2250 bp	08-SEP-2000
DEFINITION	Mus musculus nicastrin mRNA, complete cds.		
ACCESSION	AF240469		
VERSION	AF240469.1	GI:9992879	
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 2250)		

